DOCUMENT RESUME

ED 327 203 IR 053 380

TITLE	The National Coordinated Cataloging Program: An Assessment of the Pilot Project.
INSTITUTION	Council on Library Resources, Inc., Washington, D.C.
PUB DATE	Aug 90
NOTE	70p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Academic Libraries; *Bibliographic Databases; Cost Effectiveness; *Economic Factors; Higher Education; *Information Storage; Library Automation; Library Collections; *Machine Readable Cataloging; *Online Catalogs; Pilot Projects; *Standards; Tables (Data)
IDENTIFIERS	Library of Congress; *National Coordinated Cataloging Program

ABSTRACT

This report is an assessment of the National Coordinated Cataloging Program (NCCP)--a pilot project to test the idea that a set of libraries, working with the Library of Congress, can produce complete and accurate cataloging records to national standards for national distribution--and is composed of several papers and a summary report by the Bibliographic Services Study Committee (BSSC). Questions and issues addressed are: (1) the kinds of titles that should be covered by NCCP; (2) how many libraries should be in NCCP; (3) how the costs and savings of NCCP can be optimized; (4) whether the Library of Congress and the participants hold a common view of the optimum standard for a national-level quality record; (5) economic aspects of the pilot project; (6) an overview of other data considered; (7) a survey of copy cataloging practices at Association of Research Libraries (ARL) libraries; and (8) costs and cost benefits of distributed cataloging to Library of Congress standards. Tables, figures, references, and appendices are included in some chapters. (MAB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

E.B. Timmer

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

The National Coordinated Cataloging Program

An Assessment of the Pilot Project

Council on Library Resources

August 1990

Council on Library Resources

1785 Massachusetts Avenue, N.W., Suite 313

Washington, D.C. 20036

CONTENTS

	Preface	v
I	Future Directions for the National Coordinated Cataloging Program: A Review of Economic and Policy Issues Illuminated by the Pilot Project	1
II	Economic Aspects of the NCCP Pilot Project	9
III	Overview of Other Data Considered	33
IV	A Survey of Copy Cataloging Practices at ARL Libraries	47
V	Costs and Cost Benefits of Distributed Cataloging to Library of Congress Standards	61

Preface

Cataloging is what turns an accumulation of material into a library collection. Over the years, librarians have come a long way in standardizing the key elements of catalog records, thus improving prospects for sharing each other's work and, simultaneously, assisting users--especially scholars engaged in research--as they move from library to library. The rapid and imaginative adoption of computer and telecommunications technologies, as demonstrated by the bibliographic networks, has stimulated further standardization and expedited access to records and the information they represent, nationally and internationally.

The National Coordinated Cataloging Program (NCCP) is a logical next step, one in which libraries join forces to add to and expand the scope of our national bibliographic database. The precursors of NCCP--NACO (for name authorities) and CONSER (for serials)--have demonstrated that bibliographic collaboration does work. NCCP was established as a pilot project to test the idea that a set of libraries, working with the Library of Congress, can produce complete and accurate records to national standards, for national distribution. This is an essential undertaking, simply because it helps ensure that total national expenditures for bibliographic control will be kept as low as possible while maintaining high quality. Further, because libraries are facing

increased cataloging costs as they add information in new formats to their collections (e.g., databases, graphic materials) and try to respond to scholars' needs for fuller analysis of content, cost containment in every aspect of operations is essential.

This report is an assessment of the NCCP pilot project. It is composed of several papers, including a summary report by the Bibliographic Services Study Committee (BSSC) and supporting studies undertaken with Committee assistance by Paul Kantor, who, as a consultant, served as a member of the study team. There is much of interest in these reports, and they should be useful in the future development of NCCP.

BSSC was established by the Council on Library Resources to consider key issues in bibliographic control and to advise CLR on bibliographic matters. The work of the Committee is already stimulating new efforts to optimize cataloging activities from a national perspective. Members are Carol Mandel, Chair; Dorothy Gregor; and Martin Runkle. CLR itself, which has helped, in one way or another, with almost every cataloging innovation of national importance over more than thirty years, is pleased to have played a role in this study, both by funding much of the pilot project and by supporting BSSC.

Warren J. Haas

August 1990

Future Directions for the National Coordinated Cataloging Program:

A Review of Economic and Policy Issues Illuminated by the Pilot Project

Prepared for the Council on Library Resources

By the
Bibliographic Services Study Committee

Carol Mandel, Chair
Dorothy Gregor
Martin Runkle
Paul Kantor, Consultant

August 1990

Introduction

At the request of the Council on Library Resources and the Steering Committee of the National Coordinated Cataloging Program (NCCP), the Bibliographic Services Study Committee (BSSC) undertook an analysis of planning questions that could be illuminated during the initial pilot phase of NCCP. In developing its studies and presenting its analyses, the BSSC focused on general economic and policy issues related to NCCP and its role within the national bibliographic structure. The Committee assumed that evaluations of Pilot Project statistics and operational methods are best done by the Library of Congress (LC) and the participants themselves and will be reported to the Steering Committee by that group. The purpose of this report is to provide information to aid the Steering Committee in its planning for next steps and future directions of NCCP.

Because the development of a full-blown NCCP requires an operational linked systems protocol (LSP) for bibliographic records, the pilot phase of the past two years has been constrained both by start-up effects and sub-optimal telecommunications technology. It is most appropriately viewed as an exploration period, rather than as an actual pilot test. Librarians have been successfully trained, implementation questions have been hammered out, and long range planning issues have been identified and examined. Perhaps most important, this phase has fostered a highly productive dialogue between the participants and LC, a dialogue that has the potential to bring about significant positive change in U.S. cataloging practices.

The pilot phase has enabled the BSSC to address a number of questions initially raised by the Steering Committee about the optimum design and direction of the future permanent project. These questions include:

- What kinds of titles should be covered by NCCP?
- How many libraries should be in NCCP?
- How can the costs and savings of NCCP be optimized?
- Do LC and the participants hold a common view of the optimum standard for a national-level quality record?

The Steering Committee had also raised questions related to the timeliness and distribution of NCCP records. However, since the pilot project operated in a pre-LSP mode, the picture of record exchange and distribution could not be examined.

To help answer basic planning questions, BSSC has prepared an economic analysis of the NCCP pilot which is presented together with this report (Kantor, Paul B. *Economic Aspects of the NCCP Pilot Project*, 1990.) The complex relationships of the costs and savings associated with the pilot are illustrated in the analysis and help to clarify the interaction of policy decisions and economic effects.

What kind of titles should be covered by NCCP?

As noted by Avram and Wiggins,[1] the aim of NCCP is "to build a national database in which all the records are of high quality enough to be accepted into a local database or library without any modification ..., a cost-effective goal for these times of shrinking operating budgets." BSSC's study of copy cataloging costs bore out the assumption that use of an LC-quality cataloging record represents a savings for ARL libraries over use of a standard member copy record. (*Economic Aspects*, p.10; according to the study LC records are 37% less costly to use in copy cataloging by ARL libraries than are OCLC or RLIN member records.) Since libraries normally translate such savings into services for users, e.g., by increasing the production of

records added to their catalogs, availability of additional LC-quality records benefits library users. A more direct benefit in terms of enhanced retrieval may also exist, but it has not been possible to document this since wide variations in search systems and user searching strategies make it virtually impossible to isolate and compare relatively subtle differences in cataloging as they impact retrieval.

To achieve the goal of creating records for cost-effective local use, records created to LC/NCCP standard should be records that would in fact be used by other ARL libraries. One way LC has assured the creation of high-use records has been through its CIP program which captures current American imprints, a high use category for U.S. libraries. Another strategy used by LC has been to subject its cataloging priorities to review by ARL libraries. ARL-wide review and priority setting could, in the future, be applied to the development of NCCP priorities and coverage plans, increasing the probability that participants would create records needed by other libraries.

The need for a purposeful strategy is demonstrated by BSSC's study of ARL university (ARLU) library holdings for records with a 1985 imprint found in the OCLC database. When one excludes titles represented by LC records, there appears to be surprisingly little overlap of holdings--and therefore use of each other's catalog records--among ARLU libraries. Of 18,436 ARLU member copy records in one sample drawn from the database, about two-thirds were used by no other ARLU library and only 3.5% of these records were used by 10 or more ARLU libraries. While there may be additional factors that could account for this result, it does indicate that, outside the core of materials also held by the Library of Congress, high use of each other's records among ARL libraries cannot be assumed. Study and monitoring of the use of NCCP records will be important, since a key factor in the ongoing success of NCCP will be maintaining participants' contribution of relatively high use records. Likely categories for NCCP coverage might include, for example, current Western European imprints obtained through approval plans.

The OCLC sample just described excluded records also held by the Library of Congress. An LC study of use of NCCP records, which included records used both by LC and other libraries, indicates that pilot project coverage has been successful in providing records needed by other libraries. The LC sample shows an average of 12 uses of an NCCP record by ARL libraries. Even allowing for sampling error, this average is well above the two to five uses needed to achieve a "breakeven" balance between the costs of producing NCCP records and the savings to libraries of using these records. (*Economic Aspects*, p. 3)

Similarly, NCCP cataloging assignments should aim at records likely to be used by LC, since, as the economic analysis shows (*Economic Aspects* p. 4) the savings to LC of using NCCP records is considerable; these savings further can be translated by LC into the production of more LC-quality cataloging and/or can be used to off-set the costs to NCCP participants, including costs related to training. During the pilot, LC used 40% of the NCCP records created. Data gathered several months later show the LC use rate rising to 52%. This growing use rate can only be assured if particiapnts continue to add NCCP records with a high probability of use by LC. Thus NCCP assignments should continue to be coordinated with LC's priorities for current cataloging.

How many libraries should be in NCCP and which ones should they be?

Based solely on the intent to maximize the number of LC-quality records one might assume that the answer to the question of "how many" would be "as many as possible." However, as the economic analysis shows, (p. 5-8) there are a number of factors affecting the optimum size of NCCP. Because there are significant costs of NCCP contribution, both to LC and participants, the ideal number of participants will be affected by the ongoing economic balance of costs and savings, as will any particular library's decision to join the program. Even if ongoing costs are reduced, the costs of "pre-independence" training and revision place a significant burden on the

Library of Congress for every added participant. Libraries in NCCP must be those with a high probability of achieving speedy independence. If they have not already done so, pilot project participants will want to identify factors in training, operations, and the support environment conducive to gaining independence. As it does with NACO, LC will need to determine how many librarians it can train and support. As noted earlier, any library added to the program should be able to contribute records with a potential for high use. A study done by the BSSC of an OCLC database sample showed a small set of libraries creating records used as the source copy for a large fraction of the derives at ARL university libraries. However, more extensive studies are needed to confirm that this is the core group of libraries that should be included in NCCP.

How can the costs and savings of NCCP be optimized?

The economic analysis provides an overview of the costs and savings determined during the pilot phase. The cost factors were: 1) the added cost of cataloging labor at participant libraries (a median figure of 75% over ordinary original cataloging), 2) telecommunications costs, and 3) LC overhead, which dropped dramatically from $74.79 per record before a participant's independence to $4.30 for ongoing coordination. The savings factors were: 1) LC's savings in using an NCCP record for copy cataloging compared to original cataloging (amounting to $45.32 for each NCCP record used by LC or an average of $18.13 for each NCCP record created during the pilot) and 2) the savings to ARL libraries that use an LC-quality record rather than OCLC or RLIN standard member copy.

During the pilot phase, once NCCP libraries achieved independence the costs of creating NCCP records were more than outweighed by the combined savings for libraries that made use of NCCP records. An ARL library using an NCCP record rather than a member record could be expected to save $3.80. With an average of 12 ARL libraries using an NCCP record, this amounts to $45.60 per record. This is a considerable benefit of NCCP. However, it is not practical to attempt to develop a system in which these savings to libraries can be used to offset actual costs to NCCP participants. It is necessary also to look at costs and savings within the more confined universe of LC and the participating libraries.

During the pilot phase, the savings to LC were not as great as the costs to participants, even the "post-independence" costs. During the next phase of the program, it will be important to take aggressive steps toward cost reduction. While telecommunications costs should be reduced once LSP is implemented, there may be other interim strategies for operating the program. One strategy might be to increase the NCCP cataloging activity at participant libraries to maximize the use of leased lines. Another approach might be to have some participants do NCCP cataloging directly into a utility, i.e., work in the CONSER mode rather than online to LC. Since LC plans to search OCLC and RLIN for monographic copy, this model might become feasible for monographs as it currently operates for serials. However, it would also mean that NCCP participants give up searching the LC files. The cost and benefits of this "CONSER mode" of conducting NCCP merit study in the next phase of the program.

As noted previously, LC's costs for pre-independence training and revision are significant. A possible strategy for expanding NCCP might be to do so within participating libraries, where catalogers would train and assist each other, rather than to add new libraries. This would also help to increase the number of records created in relation to the cost of each telecommunications line.

The pilot project revealed at least two phases of independence. These are identified in the economic analysis as "newly independent" and "post-independence." Two NCCP participating libraries had been working online to LC for several years prior to NCCP. They achieved independence early in the project and LC's overhead costs for these two libraries were only $4.30

per record. Participants who began working online to LC as part of the NCCP pilot required far more communication with LC; LC overhead for these newly independent libraries was $35.08 per record. Because the two post-independence libraries were brought into and approached NCCP differently than the other participants, it cannot just be assumed that newly independent libraries will move through a natural transition to post-independence. During the next phase of the project LC and NCCP participants should make an effort to identify the factors ensuring the achievement of post-independence.

The greatest ongoing costs evidenced during the pilot were those expended by the participants for cataloging labor. If these costs can be sufficiently reduced, it might be possible for LC to compensate participants for their added NCCP cataloging costs out of its savings. However, the long term implications of moving from current modes of shared cataloging to what is essentially contractual cataloging need careful consideration.

Since participant costs and practices vary widely, there is a great opportunity for participants to compare their practices and identify those that are cost-effective. Equally important, however, is to continue the efforts that have begun regarding the optimization of national-level cataloging.

Do LC and the participants (and other ARL libraries) hold a common view of the optimum standard for a national-level quality record?

Discussions stimulated by the pilot project have demonstrated that the answer to this question is "not yet." While ARL librarians have agreed on what the existing standard is (i.e., AACR2, LCSH, and LC practice), they are not confident that it achieves an optimum balance between cost and quality as defined by user access. BSSC members believe that changes in accepted practice can reduce costs not only for NCCP but for all original cataloging without compromising access. The next phase of NCCP should continue to question existing practices and interpretations of standards.

Summary of issues to be considered in the next phase of NCCP implementation.

In summary, BSSC recommends that LC and the NCCP participants consider the following issues during the next phase of NCCP:

- Strategies for assuring that participants will contribute records likely to be used by other libraries.
- Providing for on-going monitoring and analysis to ensure that NCCP records are indeed used.
- Consideration of expanding NCCP contributions within participating libraries.
- Identification of factors that could lead to speedy full independence.
- Strategies for reducing pre-LSP telecommunications costs, including a possible test of a CONSER-model alternative.
- Reduction of participants' NCCP cataloging costs.

- Addressing the issue of equity inherent in NCCP, i.e., that the costs are borne by a limited number of participants while the benefits accrue to a different and larger set of institutions.

- Achievement of an optimum standard for national-level quality cataloging records.

Notes

1. Henriette Avram and Beacher Wiggins, "The National Coordinated Cataloging Program," Library Resources and Technical Services, 32:2 (April 1988).

ECONOMIC ASPECTS OF THE NCCP PILOT PROJECT

Report Prepared for the Bibliographic Services Study Committee

Paul B. Kantor
Tantalus Inc
Rev 1990 June 26

EXECUTIVE SUMMARY.

The cost components associated with NCCP 1988-89 data are summarized in Table 5. Five models are considered. In the training or pre-independence model, "**PreInd**," libraries require extensive support and coordination at LC, costing $74.79 per record. In the newly independent model, "**NewInd**" libraries are certified so that not every item is reviewed, but they maintain extensive communication with the Library of Congress. In the post-independence mode, "**PostInd**," libraries are essentially independent and require $4.30 of coordination and support per record created. Telecommunication costs are based on the FY89 experience, which was artificially expensive due to leased lines with low traffic. The fourth and fifth models include projection for "Linked System" or "**LSP**" costs, based on line charges alone. Quality Control (QC) is costed according to the LC plan for quality assurance for the NCCP libraries. FY89 experience shows that LC saves $45.00 per record that it uses (weighted average, including fringe), but that it used 40% of the records created by the NCCP libraries. The fifth model shows that, if this fraction were larger, all of the net added costs of NCCP cataloging could be **recovered** in savings at LC.

Savings are achieved at ARL University libraries (ARLU) when NCCP records are available. Those savings will cover costs for that book only if the cataloged items are held at no less than the "optimum" number of libraries listed in the last line of Table 5. A small sample drawn from the pilot project, by the LC, revealed an average of 12 derives per record created, close to the optimum for newly independent libraries. The "breakeven" number is the number of copies of NCCP titles that must be held by all the ARLU libraries if the combined savings is to cover the combined added costs of NCCP cataloging. Although the LSP Mode assures that overall savings exceed overall costs, transfer of funds among the ARLU libraries is not considered practical. We conclude that successful implementation of the NCCP concept rests on aggressive reduction of the cost items appearing in Table 5, specifically through changes in patterns of support, of telecommunication, and of cataloging practice.

Table 5. Models for the Net Added Expense

	MODEL				
Cost Category	PreInd	NewInd	PostInd	LSP Mode	Recovery
ARLU LABOR	$19.04	$19.04	$19.04	$19.04	$19.04
TELCOMM (1)	$8.21	$8.21	$8.21	NA	NA
LSP TELCOMM (2)	NA	NA	NA	$0.05	$0.05
QC (3)	$0.00	$0.37	$0.37	$0.37	$0.37
LC COORD COST	$74.79	35.08	$4.30	$4.30	$4.30
Fractn Used	40%	40%	40%	40%	53.6%
LC SAVINGS (4)	($18.13)	($18.13)	($18.13)	($18.13)	($24.29)
Net Added Exp	$84.31	$44.97	$14.19	$6.03	$0.00
ARL Saving	3.80	3.80	3.80	3.80	3.80
Optimum (5)	23.19	12.83	4.73	2.59	1.00
Breakeven (6)	33	5	2	1	1

NOTES: (1) Telecommunication costs are artificially high during the NCCP Pilot study due to the need to lease more capacity than could be used at this time.
(2) The projected cost for the LSP situation is based on line charges for communication itself which will be less than 5 cents per record. A more conservative estimate of the costs of maintaining terminals leads to a figure of approximately $2.00 per record.
(3) Quality control is based on the reported LC cost data, projected to a level of 100 records examined per year, per 1200 produced. QC at LC is based on sampling, but not at a fixed sampling percentage formula. For libraries not yet independent, the quality control process is subsumed in coordination.
(4) LC Average Savings per record Used at LC is $45.32. However, only a fraction (Fractn Used) of the records created by NCCP libraries are used at LC.
(5) The optimum is the number of libraries that must hold an item for the savings in derived cataloging from that book to cover the net added expense for that book.
(6) The breakeven point is the number of libraries that must hold a book for the total savings in derived cataloging cost from all NCCP books to cover the added expense for all NCCP books combined.

ECONOMIC ASPECTS OF THE NCCP PILOT PROJECT

1. Importance of the Economic Aspects of the NCCP Pilot Project.

Two years experience with the NCCP, in which 8 libraries in the ARL cataloged certain new materials according to the procedures used at the Library of Congress, and did so with the aid of online contact with the LC files, has established a host of "quality benefits."

In addition to reviewing these benefits, it is important to ask whether the NCCP can cover its expenses. The BSSC has supported a number of studies and surveys which clarify this issue: a study of the added cost of NCCP cataloging (compared to ordinary original cataloging); a study of the savings realized when derived cataloging is based on LC records as opposed to member records; a survey of current cataloging practices; a survey of current levels of original and derived cataloging, and record maintenance policies; and a study of the degree of overlap among the holdings of the 75 ARLU libraries which are members of the OCLC. The data from these several studies enable us to put reasonable bounds on the costs of NCCP, and on the prospects for cost recovery.

2. The Economic Balance of NCCP.

2.A. The Direct Cost Equation.

The economics of NCCP should seem to require no more than a balance between the added costs of NCCP cataloging (estimated at about $19.00 per record [KANTOR, 1989]) and the savings when a record is used at LC (estimated at about $45.00 per record [See Tables 1,4].) The savings exceed the cost and so one might imagine that LC could, in effect, pay for the added cost of new records created, and bank the difference.

Table 1. Cost Savings at Library of Congress

Category	Cost Usual	Cost Derived	Savings Direct	Fringe	Savings TOTAL
New Records	49.70	3.77	45.93	6.89	$52.82
APIF Records	49.70	9.70	40.00	6.00	$46.00
MLC Records (1)	0.00	0.00	0.00	0.00	$0.00

Note: This does not include the cost of converting records created in Dewey form into LC.
(1) LC realizes no savings when Minimal Level Cataloging (MLC) records are upgraded, since it would not do anything more to those records in any case.

Unfortunately, other factors complicate the picture. To begin with, LC must apply the Quality Control (QC) efforts to the NCCP records. This adds $4.48 per record examined. The sampling rule is essentially 64 records per year when a library's production is less than 1200 records, and 192 per year when it is above 1200. During the NCCP project 16 records were examined each month, when production permitted. In some cases this led to review of every record. Projecting production at 2300 records per NCCP library we estimate that 1 in 12 records will be examined. This adds an average of $4.48/12= $0.37 to the cost of each record created.

Far more serious is the telecommunications overhead, which, in FY89, came to $8.21 per record created. This is the cost of maintaining open lines between LC and the participant libraries throughout the working day. These two effects together increase the added cost to approximately $27.60, still substantially less than the apparent savings.

There is also a coordination cost at LC, representing the cost of personnel who communicate with and train the librarians at the NCCP libraries. For libraries that are not yet independent, this cost works out to $74.79 per record created. For the two libraries which had been fully independent for a long time, it works out to be (Table 3) substantially lower, at $4.30/record.

Table 2. Coordination Costs:Pre-Independence (Rev 90-6-7)

Activity	CostFY89	ItemsFY89	$/Item	
Descriptive	$42,736	917	$46.60	
Subject	$48,201	2814	$17.13	(1)
Marc Editing	$6,660	5107	$1.30	(1)
	$97,597		$65.04	
		Fringe	9.76	
		Total	$74.79	

Note (1). This calculation mixes several stages of independence

Table 3. Coordination Costs:Post-Independence(Rev 90-6-7)

	CostFY89	ItemsFY89
Harvard	$5,462	1,270
Chicago	$3,785	880
Total(1)	$9,247	2,150

Per NCCP Record $4.30

Note: Total Cost determined as 0.2 FTE for both libraries combined.

Cost of One	FTE	$40,206
20% of one	FTE	$8,041
	Fringe	$1,206
	Total	$9,247

(The NACO component alone is $7,770 for 2150 items = $3.61/record)

Table 3A. Coordination Costs: Newly independent				
Activity	CostFY89	ItemsFY89	$/Item	
Descriptive	$22,034	1825	$12.07	
Subject	$48,201	2814	$17.13	(1)
Marc Editing	$6,660	5107	$1.30	(1)
	$76,895		$30.51	
		Fringe	4.58	
		Total	$35.08	

Note (1). This calculation mixes several stages of independence

In Table 3A we show the condition achieved by libraries which became independent during the NCCP Pilot Project. It is apparent that their styles of operation are such that there is substantial coordination at the Library of Congress, resulting in corresponding costs per record created. It is vital to identify the factors needed to complete the transition from "Newly Independent" to "Post Independent".

Combining the coordination cost with the previous figure we find that the cost per record is $102.04 (pre-independence), $62.70 (newly independent) or $31.92 (post-independence) per record created. The post-independence figure is less than the estimated savings at LC, so there is some prospect that the savings at LC directly cover the added cost of creating an NCCP record.

This prospect is dimmed by the fact that LC realizes savings if, and only if, it holds the item corresponding to the record, and can derive its own record from the NCCP record created elsewhere. Not every NCCP record will lead to such savings.

The best estimate of the chance that an NCCP record will be usable at LC comes from the NCCP experience. (See Table 4). Through December 1989, a total of 8218 titles received NCCP cataloging, of which 3096 (38%) were also held at the LC. We round this figure to 40%.

Table 4. NCCP Experience through Dec 1989

Category	Created	Used	Save/Rec	Save TOT
New Records	6,144	1,021	$52.82	$53,929
APIF Records	1,878	1,878	$46.00	$86,388
MLC Records (1)	197	197	$0.00	$0
	8,219	3,096		$140,317
Average Fraction Used		37.67%		
Average Savings/Record Used		$45.32		

Note: This does not include the cost of converting records created in Dewey form into LC.
(1) LC realizes no savings when Minimal Level Cataloging (MLC) records are upgraded, since it would not do anything more to those records in any case.

Hence, against a net added expense of $102.04 or $31.92 per record created we can balance 40% of $45.32, or $18.13. This leaves net added expenses of $84.31 or $14.19 still uncovered. The picture changes further when we consider the artificiality of Pilot telecommunications costs.

It is difficult to project the costs to be realized when the Linked Systems Project (LSP) becomes available for bibliographic records. Various sources agree that the cost of sending individual records over the lines will be as low as five cents per record or less. The limiting factor then becomes the cost of maintaining the lines open for interactive use, and maintaining enough terminals and catalogers to keep the lines busy. We feel that $2.00 added to each record created is a reasonable estimate of these costs. If a cataloger produces 100 records per month this represents $200 per cataloger month allocated to communications costs. This is extremely conservative, since some or all of the costs of maintaining terminals would be incurred whether or not a library participates in NCCP, in which case it should not be viewed as an added expense. The more optimistic choice, including only line charges, leads to the fourth model shown in Table 5, the LSP model. The net expense per record is now reduced to $6.03.

We note that the total expense of record production, including direct labor, coordination costs, Quality Control and LSP-mode telecommunications is less than the savings at LC per record used. This means that if a high enough fraction of the records chosen for NCCP cataloging are used at the LC, the accounts can balance directly. This is shown as the [Cost] Recovery model in the fifth column of Table 5.

Table 5. Models for the Net Added Expense

	MODEL				
Cost Category	PreInd	NewInd	PostInd	LSP Mode	Recovery
ARLU LABOR	$19.04	$19.04	$19.04	$19.04	$19.04
TELCOMM (1)	$8.21	$8.21	$8.21	NA	NA
LSP TELCOMM (2)	NA	NA	NA	$0.05	$0.05
QC (3)	$0.00	$0.37	$0.37	$0.37	$0.37
LC COORD COST	$74.79	35.08	$4.30	$4.30	$4.30
Fractn Used	40%	40%	40%	40%	53.6%
LC SAVINGS (4)	($18.13)	($18.13)	($18.13)	($18.13)	($24.29)
Net Added Exp	$84.31	$44.97	$14.19	$6.03	$0.00
ARL Saving	3.80	3.80	3.80	3.80	3.80
Optimum (5)	23.19	12.83	4.73	2.59	1.00
Breakeven (6)	33	5	2	1	1

NOTES: (1) Telecommunication costs are artificially high during the NCCP Pilot study due to the need to lease more capacity than could be used at this time.
(2) The projected cost for the LSP situation is based on line charges for communication itself which will be less than 5 cents per record. A more conservative estimate of the costs of maintaining terminals leads to a figure of approximately $2.00 per record.
(3) Quality control is based on the reported LC cost data, projected to a level of 100 records examined per year, per 1200 produced. QC at LC is not based on a percentage sampling formula. For libraries not yet independent, the quality control process is subsumed in coordination.
(4) LC Average Savings per record Used at LC is $45.32. However, only a fraction (Fractn Used) of the records created by NCCP libraries are used at LC.
(5) The optimum is the number of libraries that must hold an item for the savings in derived cataloging from that book to cover the net added expense for that book.
(6) The breakeven point is the number of libraries that must hold a book for the total savings in derived cataloging cost from all NCCP books to cover the added expense for all NCCP books combined.

Are there any other savings which might cover this expense? We have studied the savings in derived cataloging, and find that it is comparable to the uncovered expense, and therefore potentially quite relevant. The implications of this are shown in the last two rows of Table 5, which are explained in detail in Section 2.B below.

2.B. Indirect Cost Benefits at ARL University Libraries.

We have found that derived cataloging from LC records is, in general, less costly than derived cataloging from "Original Member" records. Making the reasonable assumption that NCCP records will be similarly effective, we can estimate the benefit to deriving libraries. Our study has found a representative figure of $3.80 for the savings. If we divide this into the net added expense we find the number of derives needed for the savings to cover the

remaining costs.

To determine the feasibility of this mode of cost recovery we have examined the available data (on the 75 ARLU libraries in OCLC). We plotted the distribution of all titles held at ARLU libraries (including those held at LC.) We then examined the best conceivable strategy, in which the most widely held titles are cataloged first (to NCCP standards), then the next most widely held, and so on. The result is a competition between steadily growing costs, and savings that grow at a diminishing rate. The results, for each of four models [Pre-Independence, Newly Independent, Post-Independence, and LSP] are shown in Figures 1-4.

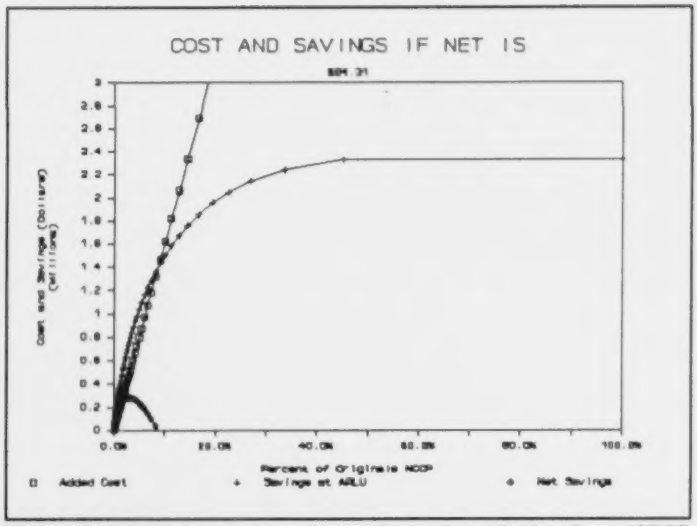

Figure 1. Achievable savings due to derived cataloging from NCCP records, shown as a function of records in the NCCP, with optimal selection. Added Cost $84.31.

Our decision to limit the analysis to benefits at the ARL University libraries is based on a number of considerations. (1) This is the community anticipated to provide the NCCP records, and so it is most natural to look for benefits within the community. Benefits flowing outside that community are thus added to the substantial list of "intangible" benefits. (2) As a practical matter, our study of the cost savings at libraries doing derived cataloging has been limited to a sample of ARL libraries, and we are not confident that the indicated cost savings would be found at other libraries.

Each graph shows, as a function of the fraction of all books which receive NCCP cataloging, the added expense of that cataloging (the straight line), the benefit to other libraries (the line that rises and becomes horizontal), and the net savings (benefit minus cost.) Tracing this last curve we note two points of interest.

The first is its maximum. At this "optimum point" only those titles held more widely than is needed to cover their own added costs are NCCP-cataloged. This leads to the maximum net savings to the ARLU libraries as a whole. The estimate of total savings depends on which estimate we use for the costs. For the pre-independence model it is $285,600. [For 75 libraries] This corresponds to NCCP cataloging for a fraction 2.9% of the titles held at the 75 ARLU libraries considered. This analysis shows the nature of the relation between cost and coverage that arises if all ARL university libraries are included in the NCCP.

Tracing the curve of net savings further we meet the "breakeven point" at which the combined savings just cover the combined costs of NCCP cataloging. In this case more widely held titles in effect subsidize the cataloging of less widely held titles. The breakeven point corresponds to 8.1% percent of all titles held at the ARLU libraries.

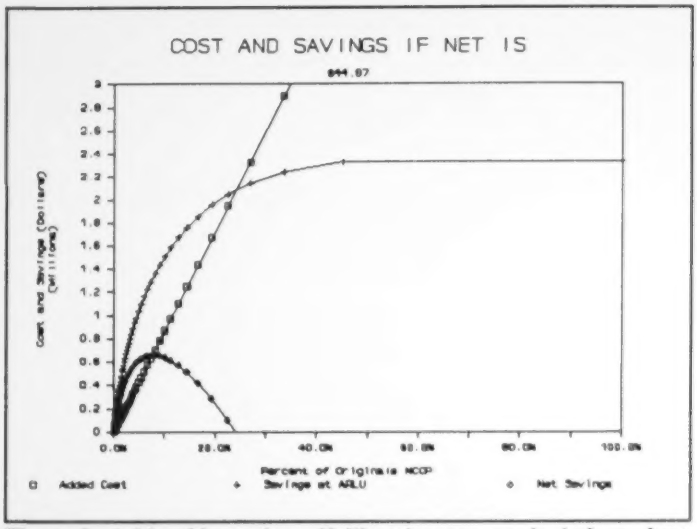

Figure 2. Achievable savings if libraries are newly independent. Added cost $44.97 per record.

We repeat the entire analysis using the newly independent model costs, in Figure 2. As the graph shows, larger savings and coverage are achieved when the net added cost of NCCP cataloging is lower ($44.97). The maximum net savings is $659,000, achieved at a coverage of 8.1% of all titles. The breakeven point moves out to represent 22.5% of all the titles held at the ARLU libraries.

We repeat the entire analysis using the more encouraging post-independence costs, in Figure 3. As the graph shows, larger savings and coverage are achieved when the net added cost of NCCP cataloging is lower ($14.19). The maximum net savings is $1,428,742, achieved at a coverage of 19.3% of all titles. The breakeven point moves out dramatically, to represent 85% of all the titles held at the ARLU libraries.

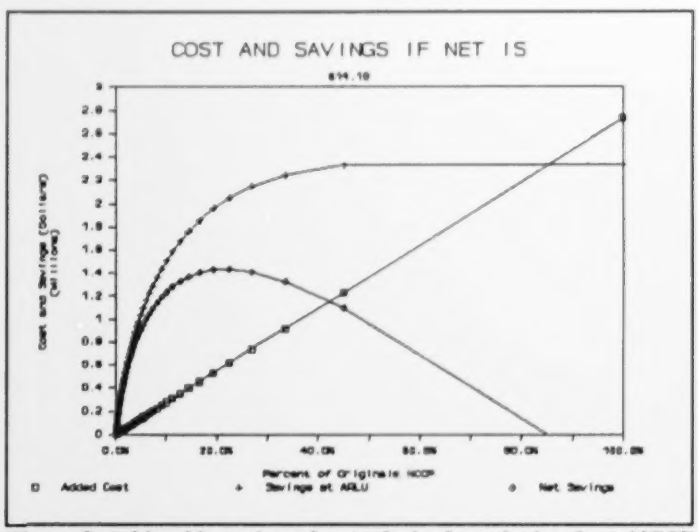

Figure 3. Achievable savings due to derived cataloging from NCCP records, shown as a function of records in the NCCP. Post-Independence Cost $14.19.

When we continue this analysis to the fourth model, using post-independence cost assumptions and the LSP line cost

estimates for telecommunications costs, we encounter a new phenomenon (see Figure 4). The net savings rises, as before, reaching a maximum of $1,854,000, representing 44.97% coverage of all the titles held by the ALRU libraries. But, as the net savings falls again, it does not reach zero before the coverage has reached 100%. This means that, if these cost figures hold, the collection of all ARLU libraries taken together could catalog all titles to the NCCP standard, and realize a net savings of just over $1,000,000. [Recall that this presumes that the savings realized at LC are also pumped into the ARL system.]

When 100% coverage can be achieved, the unreality of our assumption that the most widely held titles be cataloged first [see Appendix 2] becomes irrelevant. However, common sense suggests that the costs of transferring funds from the beneficiary libraries to the NCCP libraries might wipe out much of the savings.

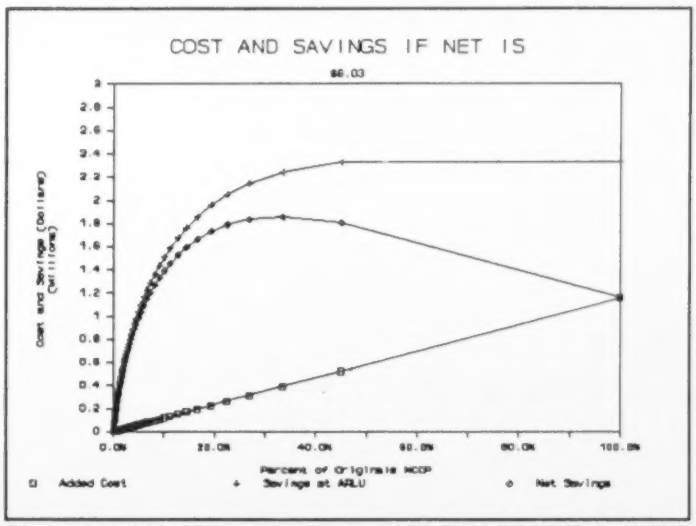

Figure 4. Achievable savings due to derived cataloging from NCCP records, shown as a function of records in NCCP. LSP Post-Independence Cost $6.03

For the moment, recovery of NCCP net added expense from the ARLU libraries does not seem to be practical. For example, to extend NCCP to all cataloging would require an enormous training and supervision effort. However, if the net added expense of NCCP cataloging were negligibly small, it could be adopted on a widespread basis, to achieve the benefits of reduced derived cataloging expense, and other quality benefits as well. So we turn to consideration of ways to decrease the net added expense.

It is perhaps worth noting that a preliminary study, by the Library of Congress, of the usage of a sample of 200 NCCP records that were at least 6 months old showed an average usage of 12 derives at ARL libraries. This represents a realized cost savings of $45.60 per record created, which is sufficient (see Table 5) to cover the net added cost for all models except the Pre-Independence model.

3. Decreasing the Net Added NCCP Expense.

Whichever estimate we use: "pre-independence," "post-independence" or "LSP-mode" we do not see present cost benefits at the Library of Congress paying for present added expense. But, all of the four components of the net added expense are subject to possible improvement. Although quality, and confidence in that quality, are essential to the qualitative

and cost benefits of NCCP cataloging, even QC costs can be lowered, per record, as production per NCCP library surpasses the level of 2300 records per year assumed in this analysis.

Telecommunications costs ($8.21 per NCCP record, or $.05 per record in LSP mode) are a fixed cost, in the sense that lines are held open throughout the working day. The cost per item will decrease as the number of items increases (until it becomes necessary to add another line.) As a temporary measure, the projected LSP costs might be attained if NCCP libraries cataloged directly into a utility, and did so without access to the LC internal files that they now are using. [This is conceptually linked to cataloging optimization, discussed below.] Our use of only the incremental line costs assumes that equipment and line costs are shared with other library programs, and utilized at full capacity. It also realizes that some or all of the cost of terminals might be incurred whether or not a library participates in NCCP, as long as it continues to catalog into some online utility.

Coordination costs at the Library of Congress (particularly for newly independent libraries at $35.08 per record) can be reduced by changes in communications patterns, by distribution of the training effort onto the NCCP libraries themselves, and by cataloging optimization.

Finally, the added direct labor cost of NCCP cataloging could be reduced as a by-product of cataloging optimization, which would identify that which is essential in the creation of records consistent with the national database, and eliminate all that is inessential. We must note, however, that this improvement could "cut in the other direction." That is, cataloging optimization could lower the usual costs at LC and thus lower the savings achieved by derived cataloging.

All in all, however, there are promising avenues for reducing the net added cost of NCCP to zero, which means that savings at the Library of Congress could, at least in principle, support this important national activity by paying for all records created, either directly, or through payment for the records that it uses, at a rate not exceeding the savings realized.

4. Summary of Data cited in the Report.

4.1 Data on Labor costs of Original and NCCP cataloging. [KANTOR 1989].

It has been established that the added cost of cataloging to NCCP standards varies widely at the 7 libraries studied. The representative median figure is a 75% increase in labor costs. Using [confidential] figures for the direct labor cost, adjusted to account for the productivity factor, and for fringe expenses we have converted this figure to a difference of **$19.04 per record created**. This is approximately confirmed by updating the cost figures given in KANTOR[1984], updated at the rate of 6% per year.

4.2 Data on the Cost Saved by Cataloging from LC Records.[KANTOR 1989]

It has been established that derived cataloging based on LC records is less expensive than derived cataloging based on member records (even where the procedures and policies are explicitly the same. The representative median figure is a 37% savings. Using the [confidential] figures for the direct labor cost, adjusted to account for the productivity factor, *and for fringe expenses we have converted this figure to a difference of* ***$3.80 per record*** derived. This is approximately confirmed by updating the copy cataloging costs reported in KANTOR 1984, and using the ratio of Member-Based to LC-based copy cataloging reported in the survey [MANDEL 1990] of ARL libraries.[61.8% LC-based; 38.2% Member based.]

4.3 Data on Telecommunications Cost [HIATT:OCT 89]

The Library of Congress reports for FY89 costs of $40,186 corresponding to the creation of 4,892 records, or $8.21/NCCP record created. Comparable data on the NACO project are dominated by the hardware expense. [Direct line charges per record come to a few cents.] The Digital Access Facility costs $1500 per month, whether it is used or not. If we project that such a device could support 7.5 catalogers it would cost $200/month per cataloger. With a production of 100 records per month per cataloger this works out to **$2.00 per record created**. Strictly speaking, however, these costs of the LSP mode might be borne by the libraries in any case, as they maintain contact with their own utilities. The added batch transfer costs are pennies.

4.4 Data on Cost Savings at the Library of Congress [HIATT: OCT 89;HIATT MAR 90].

There are several ways to approach the estimation of cost savings at the Library of Congress. One method [OCT 89] is to compare the cost of derived cataloging with the cost of original cataloging. The relevant figures are $49.70 for original cataloging without Dewey Classification, and $3.77 for copy cataloging from NCCP records. This works out to a difference of $45.93 per record derived at LC. (With Fringe added: $52.82) See Table 1. However, the savings are less when the record is already in the APIF file. The effect of this is shown in Table 1. In Table 4 we calculate the weighted average based upon experience through Dec. 1989. The **average savings per record used is $45.32**.

4.5 Data on Coordination Cost at Library of Congress.[HIATT:OCT 89; VOGEL MAY 90]

The pre-independence and newly independent estimates of coordination cost were derived by combining the results of detailed analysis done in the area of descriptive cataloging with undifferentiated cost figures for subject cataloging and MARC editing. The post-independence estimate for labor overhead at LC was based on the determination (via estimates at LC) that the two independent libraries (Harvard and Chicago) required a total of 0.2 FTE per year of support for NCCP activities. Dividing the total salary expense (including fringe and non-productive time) by the number of records produced yields a cost **per NCCP record created of $4.30**.

4.6 Comparability of Cost Data.

Data collected in the KANTOR 1989 Study were based on reporting of the number of hours spent per day at particular tasks, and the number of items processed. Data from the Library of Congress are based on estimating the percentage of total working time that is devoted to specific activities. This method accounts from 100% of paid time, and needs to be adjusted only for fringe. The method used by KANTOR 1989 does not account for 100% of time. Thus the cost figures determined in that study are "nominal costs." To adjust these for non-productive time and fringe expense we have doubled the nominal costs. This brings the two sets of cost figures into sufficient comparability that it makes sense to add and subtract them. Note that the findings [KANTOR 1989] that NCCP adds 75% to cataloging costs, and LC-based derived copy costs 37% less than member-based derived copy involve only ratios, and are unchanged by this adjustment.

5. NCCP as a lottery.

From the perspective of the libraries there is a labor expense ($19.04) which must be covered. LC realizes a savings of $45.32 per record used. So, if a library could know that enough of its records would be used by LC, it could safely catalog all of them to NCCP standards. The breakeven percentage (which can be achieved only when a library is independent) is shown as the fifth model in Table 5.

6. Acknowledgements.

This research was supported by the Council on Library Resources, and has been shaped throughout its evolution by the members of the Bibliographic Services Study Committee (Carol Mandel, Chair, Dorothy Gregor and Martin Runkle.) Data collection would have been impossible without the support and cooperation of all of the members of the NCCP Operations Group, and the staffs at 11 ARL libraries participating in the Copy Cataloging Cost study. Special thanks are due to Mark Crook, at OCLC, who developed the sample of 150,000 records, and extracted the information on holdings and overlap. Bob Hiatt and Les Vogel shared details of LC's methods for cost analysis and quality control, and clarified many potentially confusing points. Kay Guiles at LC provided the keys necessary to eliminate CIP records from the study of distribution, and to remove records that are not CIP from the CIP set. Peter McWhite, consultant to the LC, and Les Vogel and Judy Henderson clarified the details of Quality Assurance for the NCCP Pilot. Beacher Wiggins clarified the economics of telecommunications at the LC. Michael Kantor provided helpful insight into the competition between ARLU and LC for creation of the records for jointly held items.

REPORTS AND REFERENCES CITED.

CROOK, Mark. Sample of 150,000 records with imprint years 1984-1986, from the OCLC data base, broken down by originating record [CIP, other LC, non LC ARLU library] and analyzed to show the number of holdings symbols posted to each record in the sample, including only the symbols of the ARL university libraries. For technical reasons two other libraries [National Agricultural Library and New York Public Library] are included in these counts. Data provided in machine readable form, March 1990.

HIATT, Robert. OCT 89 Memo of October 1989 summarizing cost factors for FY 1989 at the Library of Congress.

HIATT, Robert. MAR 90. Memo of March 1990 summarizing the cost experience at Library of Congress through December 1989. This memo also reports the breakdown of NCCP records by originating record, and reports the fraction that were used at LC. It contains estimates of the cost of LC labor to support an independent library.

KANTOR, Paul B. [KANTOR 1984] Three Studies of the Economics of Academic Libraries, in Advances in Library Administration and Management v5(1986) Ed: McCabe, Gerard B. and Bernard Kreissman. pp221:286. JAI Press.

KANTOR, Paul B. [KANTOR 1989] Costs and Cost Benefits of Distributed Cataloging to Library of Congress Standards. Report prepared for the BSSC. Council on Library Resources. 1990 Jan.

MANDEL 1990. Survey of ARL Libraries to determine levels of Cataloging Activity. [Included as item 11 in the Overview of Other Data, bound with this report.]

VOGEL MAY 90. Fax Copy of Memorandum of May 24, 1990 for the record.

APPENDIX 1. Estimate of the Chance that a Title will be held at LC.

We have three ways to estimate the chance that an NCCP record will be usable at LC.

Two are based on the data provided by OCLC, analyzing the holdings of the LC and some (75) ARLU libraries. From this data we see that if all libraries were to participate in NCCP then since the overlap with LC is approximately 36% of the ARLU holdings, no more than 36% of the records would result in derives. [A more careful estimate lowers this to 28% (based on 1986 imprints in OCLC) because LC will get to some of the same titles first, so that the ARL University libraries will not, in fact, catalog them at all.]

The situation is more promising if a rather small number of libraries participate in the NCCP project. Books held at LC are, on average, more widely held than other titles. The 69,943 titles held at LC and ARLU correspond to an estimated 531,776 volumes held at the 75 ARLU libraries. This is an average of 7.6 copies per volume. On the other hand, the ARLU titles not held at LC have an average of 1.2 copies per title. This means that if a relatively small number of libraries participate in the NCCP, they are more likely to be picking up the titles held at LC (represented by 531,776 volumes) than those not held at LC (represented by only 150,499 volumes). On the average, a volume picked at random has a chance of 532/(532+150)= 78% of being held at the LC. This encouraging result is tempered by the fact that it represents selecting titles at random from the ARLU holdings. In reality, one must select several libraries, and if these libraries are large, they are more likely to hold a substantial number of unique titles.

The best estimate of the chance that an NCCP record will be usable at LC comes from the NCCP experience. (See Table 4). Through December 1989, a total of 8218 titles received NCCP cataloging, of which 3095 (38%) were also held at the LC.

Reviewing these three estimates, we have 28% (for total ARLU participation), 38% (from experience in the pilot project) and 78% (for an idealized small random sample). The conservative course is to use the figure based on experience, rounded to, say 40%.

Appendix 2. Cost benefits through derived cataloging.

The relative importance of the cost savings due to deriving from NCCP records, and the cost of creating those records, has been estimated for the four cases in Table 5. The estimate is based on data extracted from the OCLC database, on the overlap of holdings. Specifically, we have determined the number of titles which are held at exactly 1,2, ...,75 ARLU libraries, including both materials that were held only at ARLU libraries, and materials that were held and cataloged by the LC. This is the correct data when we want to estimate the impact, on the ARLU, of cataloging by the ARLU libraries.

The meanings of the columns in the table is as follows:

1. The number of libraries holding a title. This is based on the 75 ARLU libraries which were members of OCLC at the time of the sample and two other libraries which, for technical reasons, fell into the selected set.

2. The number of titles held at precisely that many libraries. At the head of the table appears the sampling factor. In this particular case (the sample of 1985 imprints) each title in the sample represents 5.79 titles in the entire data base.

3. The **impact** (that is, the number of derives that are facilitated) if such a title is cataloged to LC standards. When the number of holding libraries is 1, this is zero. When a title is held at two libraries, one of them can catalog it and the other can derive its record. When a title is held at three libraries, the total impact is twice the number of titles cataloged, and so on. The impact for the sample is shown in the third column.

4. The projected cost of cataloging all of the titles to the LC standard is shown in the 4th column. For example, at $84.31, each title in the sample represents a total cost of 5.79x$84.31=$488.15. In the ideal case, the most heavily held title is cataloged first. This is the last cost figure in column 4. Then, in the ideal strategy, the titles held at 73 libraries would be cataloged, bringing the cost to $976, and so on. The top number in this column is the cost if all titles were cataloged to LC standards. This column produces the straight lines in the graphs of Figures 1-3.

5. The projected savings is calculated using exactly the same principles. For example, when a title held at 75 libraries is cataloged, there are 74 derives, and a total savings of 5.9x74x$3.80=$1,628 at all the benefitting libraries. As we work up this column the savings grows ever more slowly, because we are moving to titles that are less widely held. Thus this column produces the curve that rises and becomes flat.

6. The percent of all titles cataloged at any point in this process provides the X-axis of the graph.

7. The net savings is the difference between the 5th and 4th columns. As we climb up from the bottom of the table (the ideal strategy) this number first rises, and then falls, eventually dropping again to 0. This is the breakeven point. The optimum point is the one at which the net savings is a maximum, and may be read easily from either the graph or the table.

Four versions of the table are shown here, corresponding to the pre-independence, newly independent, post-independence and LSP mode net costs.

B5TYPE4.WK1 On DATA-018 and TANDY D:\PAUL\CLR2

Table of Savings and Costs. 1985 OCLC Distribution Data.

	Sample	NCCP Cost	CumSaved			
	5.790	$84.31	$3.80			
	ARL Imp	$488.15	$22.00	Percent	NetSaves	
1	18,260	0	16,217,482	2,326,866	100.0%	(13,890,617)
2	3,821	3,821	7,303,774	2,326,866	45.0%	(4,976,908)
3	2,202	4,404	5,438,534	2,242,796	33.5%	(3,195,738)
4	1,454	4,362	4,363,617	2,145,899	26.9%	(2,217,718)
5	1,063	4,252	3,653,839	2,049,926	22.5%	(1,603,913)
6	909	4,545	3,134,931	1,956,374	19.3%	(1,178,557)
7	723	4,338	2,691,198	1,856,375	16.6%	(834,823)
8	560	3,920	2,338,262	1,760,930	14.4%	(577,332)
9	496	3,968	2,064,895	1,674,682	12.7%	(390,213)
10	400	3,600	1,822,770	1,587,378	11.2%	(235,392)
11	331	3,310	1,627,508	1,508,171	10.0%	(119,337)
12	304	3,344	1,465,929	1,435,344	9.0%	(30,585)
13	278	3,336	1,317,530	1,361,770	8.1%	44,240
14	218	2,834	1,181,823	1,288,371	7.3%	106,548
15	204	2,856	1,075,405	1,226,017	6.6%	150,612
16	205	3,075	975,822	1,163,180	6.0%	187,358
17	156	2,496	875,750	1,095,524	5.4%	219,774
18	161	2,737	799,598	1,040,607	4.9%	241,009
19	109	1,962	721,095	980,387	4.4%	259,382
20	118	2,242	667,796	937,219	4.1%	269,423
21	109	2,180	610,194	887,891	3.8%	277,697
22	90	1,890	556,985	839,926	3.4%	282,942
23	76	1,672	513,051	798,343	3.2%	285,292
24	64	1,472	475,951	761,555	2.9%	285,604
25	61	1,464	444,709	729,168	2.7%	284,459
26	45	1,125	414,932	696,957	2.6%	282,026
27	65	1,690	392,965	672,205	2.4%	279,240
28	66	1,782	361,235	635,022	2.2%	273,787
29	40	1,120	329,016	595,814	2.0%	266,798
30	47	1,363	309,490	571,172	1.9%	261,682
31	35	1,050	286,547	541,183	1.8%	254,636
32	46	1,426	269,462	518,081	1.7%	248,620
33	37	1,184	247,006	486,706	1.5%	239,700
34	36	1,188	228,945	460,656	1.4%	231,711
35	43	1,462	211,371	434,517	1.3%	223,146
36	15	525	198,380	402,351	1.2%	211,970
37	26	936	183,058	390,800	1.1%	207,741
38	19	703	170,366	370,206	1.1%	199,840
39	17	646	161,091	354,738	1.0%	193,647
40	23	897	152,792	340,525	0.9%	187,732
41	22	880	141,565	320,789	0.9%	179,224
42	17	697	130,826	301,427	0.8%	170,602
43	25	1,050	122,527	286,092	0.8%	163,565
44	16	688	110,323	262,990	0.7%	152,667
45	9	396	102,513	247,853	0.6%	145,340
46	18	810	98,119	239,140	0.6%	141,021
47	9	414	89,332	221,318	0.6%	131,986
48	17	799	84,939	212,209	0.5%	127,270
49	15	720	76,640	194,630	0.5%	117,989
50	10	490	69,318	178,788	0.4%	109,470
51	11	550	64,436	168,007	0.4%	103,571
52	10	510	59,067	155,906	0.4%	96,839
53	9	468	54,185	144,685	0.3%	90,500
54	3	159	49,792	134,388	0.3%	84,596
55	15	810	48,327	130,890	0.3%	82,563
56	8	440	41,005	113,068	0.3%	72,063
57	4	224	37,100	103,387	0.2%	66,288
58	10	570	35,147	98,459	0.2%	63,312
59	4	232	30,266	85,918	0.2%	55,652
60	5	295	28,313	80,813	0.2%	52,500
61	7	420	25,872	74,327	0.2%	48,451
62	9	549	22,455	65,082	0.1%	42,627
63	7	434	18,062	53,003	0.1%	34,941
64	5	315	14,645	43,454	0.1%	28,809
65	10	640	12,204	36,523	0.1%	24,319
66	2	130	7,322	22,442	0.0%	15,120
67	2	132	6,346	19,582	0.0%	13,236
68	4	268	5,370	16,678	0.0%	11,308
69	3	204	3,417	10,781	0.0%	7,364
70	1	69	1,953	6,293	0.0%	4,340
71	0	0	1,464	4,774	0.0%	3,310
72	1	71	1,464	4,774	0.0%	3,310
73	1	72	976	3,212	0.0%	2,236
74	0	0	488	1,628	0.0%	1,140
75	1	74	488	1,628	0.0%	1,140
76	0	0	0	0	0.0%	0
77	0	0	0	0	0.0%	0

85TYPE4.WK1 On DATA-01B and TANDY D:\PAUL\CLR2

Table of Savings and Costs. 1985 OCLC Distribution Data.

	Sample	NCCP Cost	CumSaved		
	5.790	$44.97	$3.80		
	ARL Imp	$260.38	$22.00	Percent	NetSaves
1	18,260	0	8,650,221	2,326,866	100.0% (6,323,356)
2	3,821	3,821	3,895,750	2,326,866	45.0% (1,568,885)
3	2,202	4,404	2,900,852	2,242,796	33.5% (658,056)
4	1,454	4,362	2,327,504	2,145,899	26.9% (181,605)
5	1,063	4,252	1,948,917	2,049,926	22.5% 101,010
6	909	4,545	1,672,137	1,956,374	19.3% 284,237
7	723	4,338	1,435,455	1,856,375	16.6% 420,920
8	560	3,920	1,247,202	1,760,930	14.4% 513,728
9	496	3,968	1,101,392	1,674,682	12.7% 573,290
10	400	3,600	972,245	1,587,378	11.2% 615,133
11	331	3,310	868,095	1,508,171	10.0% 640,077
12	304	3,344	781,910	1,435,344	9.0% 653,434
13	278	3,336	702,756	1,361,770	8.1% 659,014
14	218	2,834	630,371	1,288,371	7.3% 658,000
15	204	2,856	573,609	1,226,017	6.6% 652,408
16	205	3,075	520,492	1,163,180	6.0% 642,688
17	156	2,496	467,115	1,095,524	5.4% 628,409
18	161	2,737	426,496	1,040,607	4.9% 614,110
19	109	1,962	384,576	980,387	4.4% 595,811
20	118	2,242	356,195	937,219	4.1% 581,024
21	109	2,180	325,470	887,891	3.8% 562,420
22	90	1,890	297,089	839,926	3.4% 542,837
23	76	1,672	273,655	798,343	3.2% 524,687
24	64	1,472	253,867	761,555	2.9% 507,688
25	61	1,464	237,203	729,168	2.7% 491,965
26	45	1,125	221,320	696,957	2.6% 475,637
27	65	1,690	209,603	672,205	2.4% 462,602
28	66	1,782	192,678	635,022	2.2% 442,343
29	40	1,120	175,494	595,814	2.0% 420,321
30	47	1,363	165,079	571,172	1.9% 406,093
31	35	1,050	152,841	541,183	1.8% 388,342
32	46	1,426	143,728	518,081	1.7% 374,353
33	37	1,184	131,750	486,706	1.5% 354,956
34	36	1,188	122,116	460,656	1.4% 338,539
35	43	1,462	112,743	434,517	1.3% 321,775
36	15	525	101,547	402,351	1.2% 300,804
37	26	936	97,641	390,800	1.1% 293,158
38	19	703	90,871	370,206	1.1% 279,334
39	17	646	85,924	354,738	1.0% 268,814
40	23	897	81,498	340,525	0.9% 259,027
41	22	880	75,509	320,789	0.9% 245,280
42	17	697	69,781	301,427	0.8% 231,647
43	25	1,050	65,354	286,092	0.8% 220,738
44	16	688	58,845	262,990	0.7% 204,145
45	9	396	54,679	247,853	0.6% 193,174
46	18	810	52,336	239,140	0.6% 186,804
47	9	414	47,649	221,318	0.6% 173,669
48	17	799	45,305	212,209	0.5% 166,904
49	15	720	40,879	194,630	0.5% 153,751
50	10	490	36,973	178,788	0.4% 141,815
51	11	550	34,370	168,007	0.4% 133,638
52	10	510	31,506	155,906	0.4% 124,401
53	9	468	28,902	144,685	0.3% 115,783
54	3	159	26,558	134,388	0.3% 107,830
55	15	810	25,777	130,890	0.3% 105,113
56	8	440	21,872	113,068	0.3% 91,197
57	4	224	19,789	103,387	0.2% 83,599
58	10	570	18,747	98,459	0.2% 79,712
59	4	232	16,143	85,918	0.2% 69,774
60	5	295	15,102	80,813	0.2% 65,712
61	7	420	13,800	74,323	0.2% 60,523
62	9	549	11,977	65,082	0.1% 53,105
63	7	434	9,634	53,003	0.1% 43,369
64	5	315	7,811	43,454	0.1% 35,643
65	10	640	6,509	36,523	0.1% 30,014
66	2	130	3,906	22,442	0.0% 18,536
67	2	132	3,385	19,582	0.0% 16,197
68	4	268	2,864	16,678	0.0% 13,813
69	3	204	1,823	10,781	0.0% 8,958
70	1	69	1,042	6,293	0.0% 5,251
71	0	0	781	4,774	0.0% 3,993
72	1	71	781	4,774	0.0% 3,993
73	1	72	521	3,212	0.0% 2,692
74	0	0	260	1,628	0.0% 1,368
75	1	74	260	1,628	0.0% 1,368
76	0	0	0	0	0.0% 0
77	0	0	0	0	0.0% 0

B5TYPE4.WK1 On DATA-018 and TANDY D:\PAUL\CLR2

Table of Savings and Costs. 1985 OCLC Distribution Data.

		Sample	NCCP Cost	CumSaved		
		5.790	$14.19	$3.80		
		ARL Imp	$82.16	$22.00	Percent	NetSaves
1	18,260	0	2,729,523	2,326,866	100.0%	(402,657)
2	3,821	3,821	1,229,279	2,326,866	45.0%	1,097,586
3	2,202	4,404	915,346	2,242,796	33.5%	1,327,450
4	1,454	4,362	734,429	2,145,899	26.9%	1,411,470
5	1,063	4,252	614,968	2,049,926	22.5%	1,434,958
6	909	4,545	527,632	1,956,374	19.3%	1,428,742
7	723	4,338	452,949	1,856,375	16.6%	1,403,426
8	560	3,920	393,547	1,760,930	14.4%	1,367,383
9	496	3,968	347,537	1,674,682	12.7%	1,327,145
10	400	3,600	306,786	1,587,378	11.2%	1,280,592
11	331	3,310	273,922	1,508,171	10.0%	1,234,249
12	304	3,344	246,727	1,435,344	9.0%	1,188,618
13	278	3,336	221,750	1,361,770	8.1%	1,140,020
14	218	2,834	198,910	1,288,371	7.3%	1,089,462
15	204	2,856	180,999	1,226,017	6.6%	1,045,019
16	205	3,075	164,238	1,163,180	6.0%	998,942
17	156	2,496	147,395	1,095,524	5.4%	948,128
18	161	2,737	134,578	1,040,607	4.9%	906,028
19	109	1,962	121,350	980,387	4.4%	859,037
20	118	2,242	112,395	937,219	4.1%	824,824
21	109	2,180	102,700	887,891	3.8%	785,191
22	90	1,890	93,745	839,926	3.4%	746,182
23	76	1,672	86,350	798,343	3.2%	711,992
24	64	1,472	80,106	761,555	2.9%	681,449
25	61	1,464	74,848	729,168	2.7%	654,320
26	45	1,125	69,836	696,957	2.6%	627,121
27	65	1,690	66,139	672,205	2.4%	606,066
28	66	1,782	60,798	635,022	2.2%	574,223
29	40	1,120	55,376	595,814	2.0%	540,438
30	47	1,363	52,090	571,172	1.9%	519,082
31	35	1,050	48,228	541,183	1.8%	492,955
32	46	1,426	45,352	518,081	1.7%	472,729
33	37	1,184	41,573	486,706	1.5%	445,133
34	36	1,188	38,533	460,656	1.4%	422,123
35	43	1,462	35,575	434,517	1.3%	398,942
36	15	525	32,042	402,351	1.2%	370,308
37	26	936	30,810	390,800	1.1%	359,989
38	19	703	28,674	370,206	1.1%	341,532
39	17	646	27,113	354,738	1.0%	327,625
40	23	897	25,716	340,525	0.9%	314,809
41	22	880	23,826	320,789	0.9%	296,963
42	17	697	22,019	301,427	0.8%	279,408
43	25	1,050	20,622	286,092	0.8%	265,470
44	16	688	18,568	262,990	0.7%	244,422
45	9	396	17,254	247,853	0.6%	230,599
46	18	810	16,514	239,140	0.6%	222,626
47	9	414	15,035	221,318	0.6%	206,283
48	17	799	14,296	212,209	0.5%	197,913
49	15	720	12,899	194,630	0.5%	181,731
50	10	490	11,667	178,788	0.4%	167,122
51	11	550	10,845	168,007	0.4%	157,162
52	10	510	9,941	155,906	0.4%	145,965
53	9	468	9,120	144,685	0.3%	135,565
54	3	159	8,380	134,388	0.3%	126,008
55	15	810	8,134	130,890	0.3%	122,756
56	8	440	6,901	113,068	0.3%	106,167
57	4	224	6,244	103,387	0.2%	97,143
58	10	570	5,916	98,459	0.2%	92,543
59	4	232	5,094	85,918	0.2%	80,824
60	5	295	4,765	80,813	0.2%	76,048
61	7	420	4,354	74,323	0.2%	69,968
62	9	549	3,779	65,082	0.1%	61,303
63	7	434	3,040	53,003	0.1%	49,963
64	5	315	2,465	43,454	0.1%	40,989
65	10	640	2,054	36,523	0.1%	34,469
66	2	130	1,232	22,442	0.0%	21,210
67	2	132	1,068	19,582	0.0%	18,514
68	4	268	904	16,678	0.0%	15,774
69	3	204	575	10,781	0.0%	10,206
70	1	69	329	6,293	0.0%	5,964
71	0	0	246	4,774	0.0%	4,528
72	1	71	246	4,774	0.0%	4,528
73	1	72	164	3,212	0.0%	3,048
74	0	0	82	1,628	0.0%	1,546
75	1	74	82	1,628	0.0%	1,546
76	0	0	0	0	0.0%	0
77	0	0	0	0	0.0%	0

B5TYPE4.WK1 On DATA-01B and TANDY D:\PAUL\CLR2

Table of Savings and Costs. 1985 OCLC Distribution Data.

	Sample	NCCP Cost	CumSaved			
	5.790	$6.03	$3.80			
	ARL Imp	$34.91	$22.00	Percent	NetSaves	
1	18,260	0	1,159,903	2,326,866	100.0%	1,166,963
2	3,821	3,821	522,379	2,326,866	45.0%	1,804,487
3	2,202	4,404	388,974	2,242,796	33.5%	1,853,822
4	1,454	4,362	312,094	2,145,899	26.9%	1,833,805
5	1,063	4,252	261,329	2,049,926	22.5%	1,788,597
6	909	4,545	224,216	1,956,374	19.3%	1,732,158
7	723	4,338	192,479	1,856,375	16.6%	1,663,896
8	560	3,920	167,237	1,760,930	14.4%	1,593,693
9	496	3,968	147,685	1,674,682	12.7%	1,526,997
10	400	3,600	130,368	1,587,378	11.2%	1,457,011
11	331	3,310	116,402	1,508,171	10.0%	1,391,769
12	304	3,344	104,846	1,435,344	9.0%	1,330,499
13	278	3,336	94,232	1,361,770	8.1%	1,267,538
14	218	2,834	84,526	1,288,371	7.3%	1,203,845
15	204	2,856	76,915	1,226,017	6.6%	1,149,103
16	205	3,075	69,792	1,163,180	6.0%	1,093,387
17	156	2,496	62,635	1,095,524	5.4%	1,032,888
18	161	2,737	57,189	1,040,607	4.9%	983,418
19	109	1,962	51,568	980,387	4.4%	928,820
20	118	2,242	47,762	937,219	4.1%	889,457
21	109	2,180	43,642	887,891	3.8%	844,249
22	90	1,890	39,837	839,926	3.4%	800,090
23	76	1,672	36,694	798,343	3.2%	761,648
24	64	1,472	34,041	761,555	2.9%	727,514
25	61	1,464	31,806	729,168	2.7%	697,362
26	45	1,125	29,677	696,957	2.6%	667,281
27	65	1,690	28,106	672,205	2.4%	644,100
28	66	1,782	25,836	635,022	2.2%	609,186
29	40	1,120	23,532	595,814	2.0%	572,282
30	47	1,363	22,135	571,172	1.9%	549,037
31	35	1,050	20,494	541,183	1.8%	520,689
32	46	1,426	19,272	518,081	1.7%	498,809
33	37	1,184	17,666	486,706	1.5%	469,040
34	36	1,188	16,375	460,656	1.4%	444,281
35	43	1,462	15,118	434,517	1.3%	419,400
36	15	525	13,616	402,351	1.2%	388,734
37	26	936	13,093	390,800	1.1%	377,707
38	19	703	12,185	370,206	1.1%	358,021
39	17	646	11,522	354,738	1.0%	343,217
40	23	897	10,928	340,525	0.9%	329,597
41	22	880	10,125	320,789	0.9%	310,664
42	17	697	9,357	301,427	0.8%	292,071
43	25	1,050	8,763	286,092	0.8%	277,329
44	16	688	7,890	262,990	0.7%	255,099
45	9	396	7,332	247,853	0.6%	240,521
46	18	810	7,018	239,140	0.6%	232,122
47	9	414	6,389	221,318	0.6%	214,929
48	17	799	6,075	212,209	0.5%	206,134
49	15	720	5,481	194,630	0.5%	189,148
50	10	490	4,958	178,788	0.4%	173,831
51	11	550	4,609	168,007	0.4%	163,399
52	10	510	4,225	155,906	0.4%	151,682
53	9	468	3,875	144,685	0.3%	140,810
54	3	159	3,561	134,388	0.3%	130,827
55	15	810	3,456	130,890	0.3%	127,433
56	8	440	2,933	113,068	0.3%	110,136
57	4	224	2,653	103,387	0.2%	100,734
58	10	570	2,514	98,459	0.2%	95,945
59	4	232	2,165	85,918	0.2%	83,753
60	5	295	2,025	80,813	0.2%	78,788
61	7	420	1,850	74,323	0.2%	72,472
62	9	549	1,606	65,082	0.1%	63,476
63	7	434	1,292	53,003	0.1%	51,711
64	5	315	1,047	43,454	0.1%	42,407
65	10	640	873	36,523	0.1%	35,650
66	2	130	524	22,442	0.0%	21,918
67	2	132	454	19,582	0.0%	19,128
68	4	268	384	16,678	0.0%	16,293
69	3	204	244	10,781	0.0%	10,537
70	1	69	140	6,293	0.0%	6,153
71	0	0	105	4,774	0.0%	4,670
72	1	71	105	4,774	0.0%	4,670
73	1	72	70	3,212	0.0%	3,142
74	0	0	35	1,628	0.0%	1,593
75	1	74	35	1,628	0.0%	1,593
76	0	0	0	0	0.0%	0
77	0	0	0	0	0.0%	0

Appendix 3. The case of the missing Original Records.

A consistency test can be applied to the data obtained from the OCLC files, and data obtained from the survey of ARLU libraries [MANDEL 1990]. When the latter data are restricted to the OCLC libraries, both data sets yield estimates of the over-all ratio of original cataloging to derived cataloging for those libraries. It is somewhat perplexing that the estimate, from the OCLC data,[CROOK] is approximately 13 derives per original record created, while the estimate from the 1990 survey is approximately 8 derives per original. In other words, the survey indicates much more original catalog record production than is reflected in the OCLC files.

The number of records involved is not small. The ARLU libraries reported creating 349,501 original records in the most recent reporting year, and 2,650,808 derived records. If the ratio were as reported in the OCLC files, this latter number corresponds to only 203,908 original records (that is, 2,650,808 divided by 13). So there are some 145,592 records, nearly 42% of the total, not reflected in the OCLC ratio.

We can only speculate on whether this originates with differences in definitions, policies that discourage the sharing of created records, or other factors.

OVERVIEW OF OTHER DATA CONSIDERED

Prepared for the BSSC by
Paul Kantor

OVERVIEW OF OTHER DATA CONSIDERED.

Prepared by Paul Kantor

During its study of the NCCP Pilot Project, the BSSC has examined, sometimes in great detail, a mass of data about library practices and library costs. Much of this is primary data, never before collected. In the following pages we survey the classes of data considered, with notes on their key features or present disposition.

1. Preliminary survey of copy cataloging practices. This was completed and analyzed, leading to a tentative grouping of libraries by "pattern" or "type." That grouping was used to generate the random stratified sample for the ARL copy cataloging cost study. Subsequent full study of cataloging practices failed to validate that particular aspect of the study design.

2. Full study of copy cataloging practices. This resulted in a complete characterization of all the responding libraries (N = 102) in terms of how they treat the several classes of source copy: LC, CIP and Member. We found that the leading conclusions are: (1) CIP and LC source copy are generally treated the same. (2) Whichever kind of copy comes up first is used. The more complicated conclusion is that the most common pattern is to treat all types of copy uniformly, but that even this pattern is by no means dominant. Note of course that the uniform treatment at one library is generally different from that at another. A full report has been prepared on the results of this study.

Kantor, Paul B., Cherikh, Moula and Rich, Seth I. "A Survey of Copy Cataloging Practices at ARL Libraries" June 2, 1989. Tantalus Inc. Technical Report Tantalus/CT-89/1 (1989). Tantalus Inc. Cleveland Ohio 44118 (bound herein).

3. Study of the "time sequencing" of various cataloging events such as creation of CIP, Member copy completing CIP or LC replacing Member. This has been the least successful of our efforts, for several reasons. On the one hand, the survey of practices shows that nearly all ARL libraries use the first copy that becomes available to them. This invalidates our original model for the sequence of events. On the other hand, the larger on-line utility, OCLC, does not have time stamp data available. RLG does have access to such data but has reported quantitative impact-of-library data. It supplements the OCLC data discussed in item 4.

4. Impact of individual libraries. We looked into the question of which libraries produce records that have the most impact (as measured by the number of derives). This was explored for libraries in the OCLC data base, broken down by language, and results summarized in an earlier report. The effort was not brought to completion because the data were not normalized for "added cost of achieving this impact, were the library to join the NCCP." It is possible that some such ordering could be achieved by combining the results of the cost study, and the impact pictures. However, for libraries not yet in NCCP, the economics of inclusion would depend on how much it costs to do NCCP cataloging at

that particular library, something of a Catch-22.

Nonetheless, the early results from OCLC data show a strong concentration for selected foreign languages. That is, a small set of libraries create records which are the sources for a large fraction of all the derives at the ARL university libraries using OCLC. More extensive studies would be needed to determine whether this is true for the entire set of ARL university libraries, and to determine whether the same libraries are in the core from year to year. If so, then they are candidates for early inclusion in the NCCP, pending information on the costs of including them. In Figure 1 we show the cumulated distribution for French language books, based on the OCLC 0.1% sample of all records. In Figure 2 we show the corresponding distribution for all non-English language books.

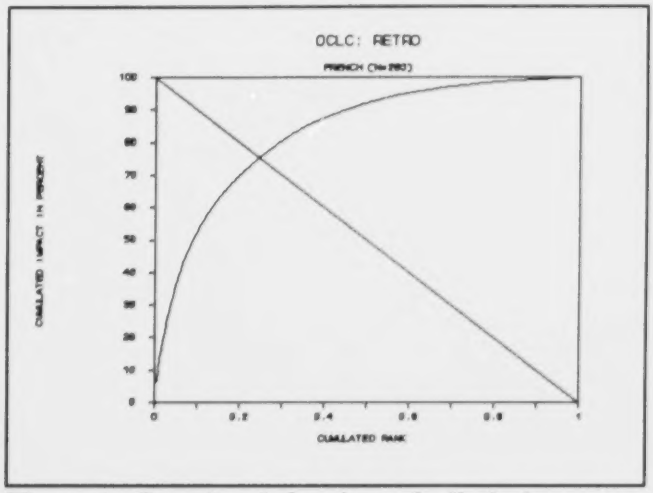

Figure 1. Cumulated fraction of all derives, as a function of the number of libraries included.

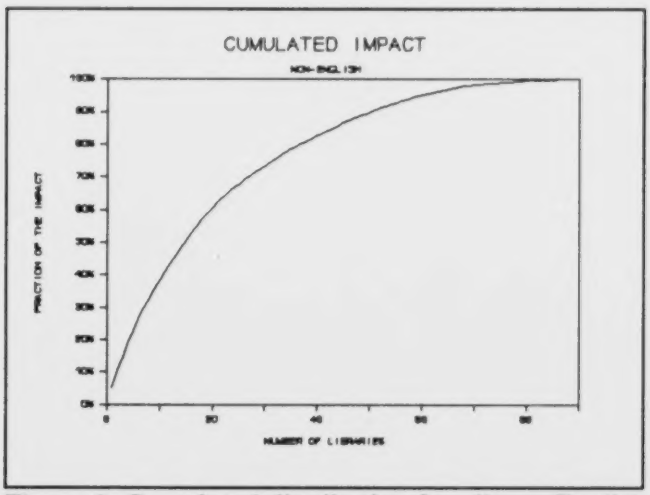

Figure 2. Cumulated distribution for all non-English titles in the sample.

Kantor, Paul B. "Second Report on OCLC Concentration of Impact Studies." May 23, 1989. Tantalus Inc.Progress Report, for distribution only to the BSSC.

5. Cost studies. We completed the study of the added cost of NCCP cataloging at the participating libraries (except Yale). This gave the fairly clean result that the percentage increase varies a lot, with a median figure of about 75% increase. Figures at one of the most experienced libraries are lower. This has also been worked into a (more rough and less certain) average value of about $19.00 for the added cost of NCCP.

Kantor, Paul B. "Cost and Cost Benefits of Distributed Cataloging to Library of Congress Standards". Tantalus Inc. Technical Report Tantalus/CT-89/6. January 22, 1990. This report is bound herein.

6. We completed a study of the added cost of deriving from member rather than LC/CIP. This is somewhat confounded by the fact that some libraries do authority work contingent on the LC processing, which appears as an increase in cost, but is really an additional benefit. Removing the obvious example of this, we derived a figure of about 35% saved in working from LC/CIP. As in the preceding case we have also estimated an absolute dollar figure of about $3.80 saved by deriving from LC/CIP.

Tantalus Inc. Technical Report Tantalus/CT-89/6. January 22, 1990.

7. Cost studies at the Library of Congress. The Library of Congress produced a stream of data whose value increased dramatically as key issues and needs of the project became clearer. These include: actual and projected telecommunication costs, and cost of LC coordination broken down into pre-independence, newly independent, and post independence of the libraries. These data are cited as appropriate in the BSSC report on economic aspects of the NCCP. Otherwise, they are treated as proprietary to the Library of Congress, and have not been reported elsewhere.

8. Impact studies at the LC. The Library of Congress has produced an estimate of the impact of NCCP titles at LC, by actual count. This may provide a lower limit, since they may turn up other books in their work flow later. They have also done a sampling study, examining both the OCLC and the RLG data bases, to estimate the impact of NCCP on the ARL libraries. The latest figures (Spring 1990) are 40-50% for impact at LC, and an average of about 12 derives at ARL university libraries per NCCP record. These figures have been used, as appropriate in the BSSC report on economic aspects of the NCCP. Otherwise, they are treated as proprietary to the Library of Congress, and have not been reported elsewhere.

9. Overlap of holdings at ARLU libraries. This turned out to be an important factor in resolving the apparently paradoxical estimates that NCCP works in the small and does not pay for itself in the large. Data were gathered at OCLC for three imprint years, and broken down into CIP, LC, and not LC (i.e. member original). The last two categories were combined to study the overlap, among ARL libraries, of the holdings of books requiring original cataloging. These have been incorporated in the BSSC report on economic aspects of the NCCP.

There is an additional observation that can be drawn from these studies, on the relative

impact of each of the three types of source copy on derived cataloging generally. We illustrate these here with cumulated graphs showing the impact of each type of cataloging, measured in the number of derives that take place at ARL university libraries. As with the other OCLC data, this is for a specific imprint year (1985), represents a sample (sample results have been multiplied by 5.79 to provide estimated totals), and applies only to those ARL university libraries which are members of OCLC. The detailed tables supporting Figure 3 are included as Exhibit 1 at the end of this report. It is likely that, although absolute figures will change, the ratio of 95:50:15 for the relative contributions is true also for the entire set of ARL university libraries.

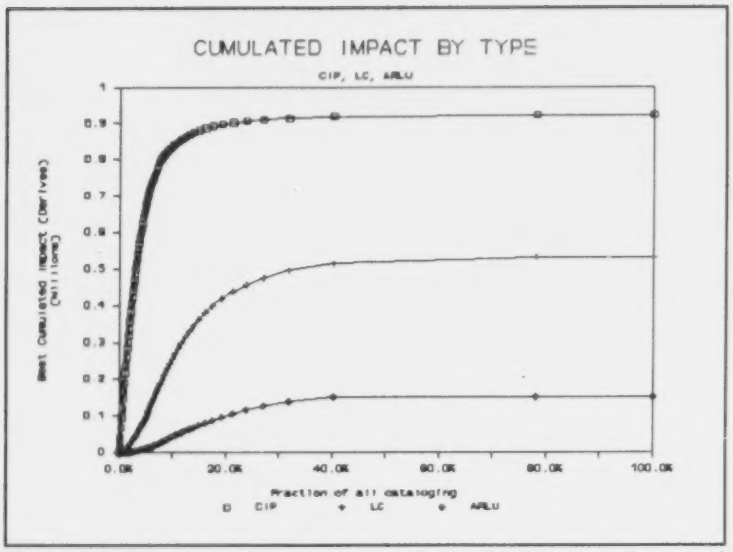

Figure 3. Impact of each type of cataloging on the total derived cataloging at ARLU libraries.

Figure 3 shows the cumulated impact (measured in total derives) as a function of the fraction of all original cataloging that is completed. These curves assume that the most widely held titles are cataloged first, as in our economic models. The most important observation is that, at 100% cataloging, the impact of CIP is clearly dominant, accounting for nearly one million derives. Other LC cataloging is also an important contributor, while ARL universities, as a group, have lower impact. This reflects, in part, the fact that the ARL university libraries must catalog their own unique holdings, which contribute no impact at all on the derived cataloging.

Figure 3 reveals that the CIP impact in fact rises very rapidly, because a substantial number of titles are held at more than 65 libraries. As noted, it contributes an estimated 921,000 derives at this set of libraries. The second most important source is Library of Congress non-CIP cataloging. Its impact rises more slowly, but finally contributes about 531,000 derives. The third source, ARL member cataloging rises most slowly, with relatively few titles held at more than 30 libraries. Overall it contributes about 150,000 derives at this set of libraries.

The data on which this analysis is based are a random sample drawn from OCLC records in such a way as to produce a sample of approximately 50,000 records. The data have been, other than their analysis for the reports of the BSSC, treated as proprietary to OCLC.

10. Studies of "randomized or statistical models for effects of NCCP." Researchers at Tantalus put considerable effort into developing spreadsheet models to estimate the impact of expanding NCCP in a less than optimal way. Several spreadsheets were developed using maximum likelihood calculations, hypergeometric distributions, and other tools of the statistician's trade. In the final analysis it proved impossible to project from sample to the entire ARL frame in a satisfactory way. At the same time, the detailed economic results have established that rational (indeed, optimal) expansion of the NCCP is vital to its success.

The analysis of the "reflected effect" falls into this same category. The reflected effect assumed that LC would apply savings to the cataloging of additional titles selected at random. This gives rise to a complex random competition between the ARL and the LC to catalog the jointly held titles.

11. The survey of cataloging volume at ARL university libraries. This has provided some data which can be used to check the endpoint of our cost models against the OCLC-based models. There are some unresolved inconsistencies still. In any case, these data are valuable in their own right, and can be massaged in a number of ways. We illustrate this by presenting a few summary tables. The rank order tables from which these results are derived is included as Exhibit 2 at the end of this report. The reader is warned that these data have not been reconfirmed with the participating libraries, and in some cases are "suspicious looking". The overall trend of the data, however, appears sensible.

Table 1. SUMMARY TABLE OF CATALOGING ACTIVITY (N=89 Libraries).

	MonTit	Full Orig	LC Copy	AllCopy	MembrBsd
DECILE	66,048	9,559	41,556	54,471	21,912
QUART	48,937	5,300	26,701	41,601	13,941
MEDIAN	30,844	2,699	18,218	26,486	6,633
AVERAG	37,651	4,033	20,923	29,924	10,045
TOTAL	3,350,926	358,940	1,799,352	2,663,213	863,861

Only ten percent of the libraries are at or above the DECILE value. We see that the distributions are skewed, with the average in every case higher than the mean. Even at the average library only 4,000 titles receive original cataloging per year, representing the work of approximately four full time catalogers. More than seven times that many titles are copy cataloged. Member based copy represents about one third of all copy cataloging. As mentioned in the report on economics, this ratio is not consistent with the ratio that is derived from the OCLC data. It shows a much larger absolute quantity (and hence, proportion) of original cataloging. It may be that the data analyzed in this table include many materials uniquely held and not suitable for inclusion in OCLC.

No other report of these data and their analysis has been issued.

DIRECTIONS FOR FURTHER RESEARCH.

Three of the directions which are opened above seem worth continuing at some time.

1. Full ARL overlap studies. The cumulated impact studies, which are very useful for understanding the cost savings potential and the expansion path of the NCCP, should be extended to include those ARL university libraries not in OCLC. This can be done, at least in principle, by searching the same sample of titles drawn at OCLC, in the RLIN database, and preparing a complementary sample drawn at random from the RLIN database.

2. Relations between cataloging practices and levels of activity. The study of cataloging practices, which did show some correlation between practice and Gross Volumes Added (an ARL statistic) might reveal more valuable insights if it is considered in terms of the detailed data on levels of cataloging reported in the survey of cataloging volume.

3. Extended impact studies. We have found, not surprisingly, that a few libraries produce records responsible for a large fraction of the derived records. This result could guide the selection of new libraries to be added to the NCCP. For this to happen the study should be expanded to include the impact on non-OCLC ARL university libraries, and should be done for several samples, to establish that membership in the core is approximately stable. Once this is done, estimates could be made of the impact of those libraries on the Library of Congress, both in original cataloging avoided, and in training and coordination costs. Finally, the cost, to that library, of cataloging in the NCCP would have to be estimated. Thus impact on the ARL copy cataloging could serve as one positive indicator for inclusion in the NCCP.

EXHIBITS

Exhibit 1. Cumulated data on the impact of CIP, LC and ARL cataloging measured in derived cataloging.

Col.	Significance
1	Number of ARL university libraries holding the title
2	Number of records, in the sample, having that number of holdings, and produced by CIP
3	Number of records, in the sample, having that number of holdings, and produced by LC
4	Number of records, in the sample, having that number of holdings, and produced by ARL university libraries.
5-7	The impact, measured in derives, of each type of production.
8-10	The cumulated impact, cumulated from the "bottom up". This means that the records having the greatest impact are considered first.
11	Percent of all titles in the sample, cumulated bottom up.

Exhibit 2.

The data reported by 89 ARL university libraries are reported in rank order, from the largest value to the smallest value, for each of several variables. Three of the libraries reported that they regularly update catalog records on the basis of updated bibliographic records received from the Library of Congress.

The rows representing the top 10% (DECILE), the top 25% (QUART), the MEDIAN (50%), three quarters and 90% point are labeled.

Note that (1) these data have not been reconfirmed with the libraries and (2) the data elements in the same row do not, in general, refer to the same library.

The columns correspond to:
Monographic titles cataloged
Full original cataloging
LC (or CIP)- based copy cataloging
All copy cataloging
Member- Based copy cataloging
Full original cataloging as a percentage of titles cataloged.

[Worksheet: Data-18\85TYPE4.WK1. Range COMP. 06/26/90]

Exhibit 1. Comparative Impact Tables: 1985 OCLC Sample Data.

Holdings	n(k) CIP	LC only	ARLU	Impact CIP	LC only	ARLU	Cumulated Impact (Scaled by sampling f CIP	LC only	ARLU	FrctnHold
0	1,129	9,673	102	0	0	0	921,722	531,777	150,499	100.00%
1	661	3,025	15,235	661	3,025	0	921,722	531,777	150,499	78.16%
2	367	1,691	2,130	734	3,382	2,130	917,894	514,262	150,499	40.26%
3	199	1,165	1,037	597	3,495	2,074	913,645	494,680	138,167	31.87%
4	170	832	622	680	3,328	1,866	910,188	474,444	126,158	27.06%
5	167	586	477	835	2,930	1,908	906,251	455,175	115,354	23.81%
6	132	594	315	792	3,564	1,575	901,416	438,210	104,307	21.35%
7	135	481	242	945	3,367	1,452	896,830	417,575	95,188	19.26%
8	130	375	185	1,040	3,000	1,295	891,359	398,080	86,781	17.54%
9	107	357	139	963	3,213	1,112	885,337	380,710	79,282	16.16%
10	98	292	108	980	2,920	972	879,762	362,107	72,844	14.95%
11	86	247	84	946	2,717	840	874,087	345,200	67,216	13.96%
12	84	231	73	1,008	2,772	803	868,610	329,468	62,353	13.12%
13	79	218	60	1,027	2,834	720	862,774	313,418	57,703	12.34%
14	79	166	52	1,106	2,324	676	856,827	297,010	53,534	11.63%
15	70	168	36	1,050	2,520	504	850,424	283,554	49,620	11.03%
16	60	161	44	960	2,576	660	844,344	268,963	46,702	10.49%
17	78	119	37	1,326	2,023	592	838,786	254,048	42,881	9.95%
18	61	125	36	1,098	2,250	612	831,108	242,335	39,453	9.49%
19	61	86	23	1,159	1,634	414	824,751	229,307	35,910	9.04%
20	40	81	37	800	1,620	703	818,040	219,846	33,513	8.70%
21	57	100	9	1,197	2,100	180	813,408	210,467	29,442	8.38%
22	46	70	20	1,012	1,540	420	806,478	198,308	28,400	8.05%
23	61	62	14	1,403	1,426	308	800,618	189,391	25,968	7.78%
24	61	50	14	1,464	1,200	322	792,495	181,134	24,185	7.51%
25	59	51	10	1,475	1,275	240	784,018	174,186	22,320	7.25%
26	63	41	4	1,638	1,066	100	775,478	166,804	20,931	7.01%
27	67	54	11	1,809	1,458	286	765,994	160,632	20,352	6.80%
28	51	52	14	1,428	1,456	378	755,520	152,190	18,696	6.53%
29	50	32	8	1,450	928	224	747,252	143,760	16,507	6.30%
30	34	43	4	1,020	1,290	116	738,856	138,387	15,210	6.12%
31	63	32	3	1,953	992	90	732,950	130,918	14,539	5.96%
32	45	43	3	1,440	1,376	93	721,642	125,174	14,018	5.76%
33	44	35	2	1,452	1,155	64	713,305	117,207	13,479	5.58%
34	41	30	6	1,394	1,020	198	704,898	110,520	13,109	5.42%
35	45	38	5	1,575	1,330	170	696,826	104,614	11,962	5.26%
36	51	11	4	1,836	396	140	687,707	96,913	10,978	5.09%
37	52	24	2	1,924	888	72	677,077	94,620	10,167	4.95%
38	60	19	0	2,280	722	0	665,937	89,479	9,750	4.80%
39	62	12	5	2,418	468	190	652,736	85,298	9,750	4.64%
40	66	20	3	2,640	800	117	638,735	82,589	8,650	4.48%
41	55	21	1	2,255	861	40	623,450	77,957	7,973	4.30%
42	50	14	3	2,100	588	123	610,393	72,971	7,741	4.15%
43	56	21	4	2,408	903	168	598,234	69,567	7,029	4.01%
44	57	15	1	2,508	660	43	584,292	64,338	6,056	3.85%
45	56	9	0	2,520	405	0	569,771	60,517	5,807	3.71%
46	49	14	4	2,254	644	180	555,180	58,172	5,807	3.58%
47	52	9	0	2,444	423	0	542,129	54,443	4,765	3.44%
48	44	13	4	2,112	624	188	527,979	51,994	4,765	3.32%
49	49	13	2	2,401	637	96	515,750	48,381	3,677	3.20%
50	57	8	2	2,850	400	98	501,848	44,693	3,121	3.07%
51	57	11	0	2,907	561	0	485,347	42,377	2,553	2.93%
52	44	8	2	2,288	416	102	468,515	39,129	2,553	2.80%
53	49	8	1	2,597	424	52	455,268	36,720	1,963	2.69%
54	62	2	1	3,348	108	53	440,231	34,265	1,662	2.57%

(continued)

Holdings	n(k) CIP	LC	Impact ARLU	CIP	LC	ARLU	Cumulated Impact (Scaled by sampling f CIP	LC only	ARLU	FrctnHold
55	43	14	1	2,365	770	54	420,846	33,640	1,355	2.44%
56	72	7	1	4,032	392	55	407,153	29,182	1,042	2.33%
57	46	4	0	2,622	228	0	383,808	26,912	724	2.17%
58	38	10	0	2,204	580	0	368,626	25,592	724	2.07%
59	52	4	0	3,068	236	0	355,865	22,234	724	1.97%
60	53	5	0	3,180	300	0	338,101	20,867	724	1.86%
61	53	7	0	3,233	427	0	319,689	19,130	724	1.74%
62	48	9	0	2,976	558	0	300,970	16,658	724	1.62%
63	61	6	1	3,843	378	62	283,739	13,427	724	1.51%
64	59	4	1	3,776	256	63	261,488	11,238	365	1.37%
65	61	10	0	3,965	650	0	239,625	9,756	0	1.24%
66	67	2	0	4,422	132	0	216,668	5,993	0	1.10%
67	60	2	0	4,020	134	0	191,064	5,228	0	0.96%
68	53	4	0	3,604	272	0	167,788	4,453	0	0.84%
69	68	3	0	4,692	207	0	146,921	2,878	0	0.73%
70	62	1	0	4,340	70	0	119,755	1,679	0	0.58%
71	64	0	0	4,544	0	0	94,626	1,274	0	0.46%
72	60	1	0	4,320	72	0	68,316	1,274	0	0.33%
73	42	1	0	3,066	73	0	43,303	857	0	0.21%
74	30	0	0	2,220	0	0	25,551	434	0	0.12%
75	17	1	0	1,275	75	0	12,697	434	0	0.06%
76	6	0	0	456	0	0	5,315	0	0	0.02%
77	6	0	0	462	0	0	2,675	0	0	0.01%
78	0	0	0	0	0	0	0	0	0	0.00%
79	0	0	0	0	0	0	0	0	0	0.00%
80	0	0	0	0	0	0	0	0	0	0.00%
81	0	0	0	0	0	0	0	0	0	0.00%
82	0	0	0	0	0	0	0	0	0	0.00%
TOTALS	6,929	21,753	21,244 (Sample Only)							

Exhibit 2. Sorted Data on Cataloging Activity (N=89)

Rank	MonTit	Full Orig	LC Copy	AllCopy	MembrBsd	Full:Titls
1	122,185	25,859	55,476	89,098	64,314	91.1%
2	100,378	22,643	54,471	84,278	44,915	41.2%
3	92,913	16,075	53,808	78,651	44,003	30.8%
4	88,820	14,007	52,655	77,388	43,575	26.2%
5	74,301	12,989	47,747	70,389	29,625	22.6%
6	74,017	11,432	45,719	65,010	26,256	21.6%
7	73,458	11,257	45,590	61,381	23,002	21.2%
8	73,416	10,000	44,183	58,061	22,642	21.1%
DECILE	66,048	9,559	41,556	54,471	21,912	19.0%
10	65,858	9,531	40,703	53,808	21,058	18.6%
11	62,400	9,510	38,221	52,655	19,933	18.5%
12	60,822	9,005	37,624	50,773	18,538	17.9%
13	60,457	8,592	36,809	50,773	18,100	15.1%
14	58,595	8,476	35,125	49,578	17,773	15.0%
15	57,738	7,357	34,648	49,396	16,651	13.8%
16	55,374	7,336	34,050	48,535	16,445	13.8%
17	52,950	6,993	34,049	45,719	15,565	13.3%
18	52,712	6,841	30,321	45,590	15,509	13.0%
19	51,232	6,364	28,436	45,196	15,128	12.9%
20	49,961	5,778	27,760	43,521	14,315	12.9%
21	49,854	5,434	27,477	41,701	14,062	12.9%
QUART	48,937	5,300	26,701	41,601	13,941	12.8%
23	48,501	4,777	26,394	39,917	13,149	12.8%
24	47,650	4,716	26,092	38,353	11,588	12.2%
25	47,343	4,359	25,748	36,809	11,357	11.8%
26	45,194	4,200	24,537	35,650	11,147	11.6%
27	44,237	4,124	24,471	35,650	11,123	11.4%
28	42,719	4,005	22,900	35,262	10,951	11.4%
29	41,687	3,996	21,908	34,050	10,709	11.1%
30	41,525	3,876	21,277	33,150	10,573	11.0%
31	40,950	3,770	21,257	32,429	10,422	10.6%
32	38,403	3,663	20,969	32,011	10,179	10.5%
33	38,250	3,587	20,789	31,986	9,616	10.4%
34	37,753	3,547	20,550	30,995	9,256	10.3%
35	37,670	3,372	20,432	30,166	9,241	10.0%
36	37,424	3,330	20,423	30,146	9,225	9.9%
37	36,620	3,250	20,390	30,051	9,217	9.7%
38	35,698	3,069	20,165	29,700	8,525	9.2%
39	34,433	2,929	20,085	29,421	8,050	9.0%
40	34,425	2,925	19,984	28,358	7,968	8.6%
41	32,532	2,911	19,967	27,651	7,707	8.6%
42	32,418	2,753	19,872	27,032	7,472	8.4%
43	31,906	2,713	19,478	26,743	7,304	8.0%
44	31,198	2,713	18,835	26,699	6,761	7.7%
MEDIAN	30,844	2,699	18,218	26,486	6,633	7.5%
46	30,328	2,632	17,700	25,853	6,521	7.5%
47	29,815	2,396	17,676	25,024	6,054	7.4%
48	29,729	2,026	17,334	24,989	5,970	7.2%
49	28,692	2,011	17,182	24,887	5,932	7.0%
50	28,392	1,983	17,157	24,832	5,871	6.6%
51	27,558	1,939	17,135	24,537	5,818	6.4%
52	27,385	1,782	17,112	24,486	5,728	6.4%

(continued)

Rank	MonTit	Full Orig	LC Copy	AllCopy	MembrBsd	Full:Titls
53	26,454	1,708	16,974	24,471	5,568	6.2%
54	26,167	1,638	16,238	24,461	5,284	6.2%
55	26,043	1,600	16,047	24,451	4,941	6.1%
56	26,000	1,506	15,444	22,871	4,930	6.0%
57	25,701	1,499	14,940	22,301	4,884	5.9%
58	25,173	1,490	14,829	22,244	4,592	5.8%
59	25,044	1,479	14,567	22,017	4,536	5.5%
60	24,674	1,370	13,881	21,331	4,308	5.2%
61	24,432	1,298	13,145	21,266	4,196	5.0%
62	24,302	1,167	13,050	20,818	3,932	5.0%
63	24,148	1,143	12,719	20,758	3,799	4.7%
64	23,914	1,103	12,665	20,001	3,662	4.4%
65	23,222	1,059	11,904	19,240	3,630	4.4%
66	23,092	1,053	11,600	19,190	3,092	4.4%
3QUART	22,100	1,043	11,593	18,921	2,735	4.2%
68	21,920	986	11,483	18,713	2,443	4.0%
69	21,861	921	11,000	18,179	0	4.0%
70	21,767	872	10,760	17,676	0	3.7%
71	21,371	860	10,383	17,257	unk	3.2%
72	21,179	814	10,003	17,157	n/a	3.1%
73	21,017	796	9,251	14,933	n/a	3.0%
74	20,559	793	8,182	14,919	n/a	2.9%
75	20,249	686	7,817	14,535	n/a	2.7%
76	20,082	665	7,681	11,989	combined	2.6%
77	19,823	589	6,861	11,271	combined	2.6%
78	19,750	588	5,339	10,260	combined	2.5%
79	19,696	558	1,134	9,953	combined	2.3%
9DEC	19,168	540	696	1,134	combined	2.3%
81	16,145	526	unk	0	combined	2.2%
82	15,986	509	n/a	0	combined	2.2%
83	15,791	501	n/a	0	combined	1.9%
84	15,276	462	n/a	0	combined	1.6%
85	15,211	316	n/a	0	combined	1.4%
86	13,330	223		0		1.2%
87	13,197	188		0		0.9%
88	12,577	unk		0		0.0%
89	10,800	n/a		0		0.0%

A SURVEY OF COPY CATALOGING PRACTICES AT ARL LIBRARIES

A SURVEY OF COPY CATALOGING PRACTICES AT ARL LIBRARIES

Paul B. Kantor and Moula Cherikh,
with the assistance of Seth I. Rich

Tantalus Inc.
June 2, 1989

Abstract

We have analyzed a survey of copy cataloging practices at ARL libraries to search for any dominant patterns of copy cataloging practice and/or staffing. Although individual procedures show some strong concentration of behavior, when a range of copy cataloging activities is examined, these concentrations dissolve in a welter of idiosyncratic patterns. We do find 11 libraries which process Member, LC and CIP copy according to the same rules and with the same personnel. Beyond that, the most common pattern is to treat only Member copy differently. Together, these two patterns of behavior (no distinction and only Member different) are found at just over half (54) of the 102 libraries reporting in the survey.

1. Overview and Recommendations

Unlike the happy families of Tolstoy's world, the happy libraries of the Association of Research Libraries are each happy in some unique way. This report explores the substantial variation in several aspects of copy cataloging.

The Bibliographic Services Study Committee surveyed copy cataloging practices, as a first step in selecting a sample of libraries for a study of copy cataloging costs. The results were somewhat surprising, in that no clear patterns of behavior were found to dominate across a substantial number of libraries. At that time, a preliminary analysis was made, which identified the "most popular patterns of similarity of treatment" and the "most popular patterns of dissimilarity of treatment," and libraries were scored according to the excess of similarity over dissimilarity. The analysis was not easily followed, and left the unsatisfied feeling that there must be some patterns here, which we were just not seeing.

With this in mind, the BSSC retained Tantalus Inc to carry out a two-stage investigation of the problem. The first stage was to be a more detailed look at the issue, with a second step to be taken only if the results of the first step seemed to warrant it. The present report is the conclusion of the first step. The results, summarized very briefly (more details are given in the body of the report) are as follóws:

A. Results for each aspect of copy cataloging may be scored on a 5 point scale (5 = LC, CIP and Member copy treated the same, 4 = Only Member copy treated differently, 3 = Only LC copy treated differently, 2 = Only CIP Copy treated differently, 1 = All three classes treated differently from each other). Each library can be represented by a set of five numbers, showing how the several aspects:

[1] = Waiting for better copy,
[2] = Verification of call numbers,
[3] = Revision practice,
[4] = Headings verification and authority work,
[5] = Staff involved,

are treated. For example, a library represented by the numbers 53412 would be one at which the policy on waiting was the same for all, verification of call numbers was different only for LC copy, revision practice was different only for Member copy, headings verification was different for all three types of copy, and the staff involved are different only for CIP.

B. In this concise language, the only significant concentration is the pattern 55555. (No distinctions, for all 5 aspects.) This pattern occurs in a total of 11 cases. This is more than the expected number (4) if the individual cases were chosen at random subject to the frequency Table III.

C. Beyond that, the most common deviation from a "5" is, as was expected, a "4." The total number of libraries for which each aspect is either handled as a "5" or a "4" is over 50%. This is approximately the expected number if the individual cases were chosen at random subject to the frequency Table III. However, this large set includes a great many variations on the theme. (See Section 3 below).

D. It had been speculated that staffing patterns and/or the maturity of the library's in-house on-line system might clear up the mess, by accounting for the variations in pattern. Analysis does not support this hope. Some correlation with library size was found in staffing and policy patterns. (See Section 6.)

E. The results of our analysis are summarized in Table I. Our recommendation is that no further resources be invested in this inquiry.

TABLE I. Summary Data on Patterns of Copy Cataloging

Activity	LC-CIP-Member Same	Only Member Different
Use at once?	79	16
Verify call number?	59	40
Revision practice	54	34
Verification/Authority	37	44
Same staff used	43	37

Professional staff do copy cataloging of all types at 12 libraries.
Professional staff do only Member copy cataloging at 24 libraries.

Number of libraries treating all types of copy the same in all aspects: 11
Number of libraries treating only Member copy different in:

Exactly one aspect:	14
Exactly two aspects:	11
Exactly three aspects:	10
Exactly four aspects:	5
Every aspect:	3
TOTAL	54

F. In the remainder of this report we elaborate the methods and results summarized above.

2. Summary of the Data.

Examining the survey instrument (Appendix I) it seems that, in nearly all cases, a library should check only one box in any row of any table. In fact, there were many

exceptions to this rule, which complicated the preliminary data analysis enormously. In any case, our report begins with a summary of the total number of checks appearing in each box (Table II). We see that some patterns are quite clear: LC copy is always accepted, for example. But, even though some of the other rows contain large numbers, our detailed analysis reveals that "it was not the same libraries" in each question, or even in each row.

We note, for the record, the on-line systems in use. Several libraries reported the use of more than one on-line, or of a specific in-house system. Those data are not summarized here.

On-Line System Used:

OCLC	RLIN	Neither
78	32	11

The summary of responses is laid out to correspond to the questions of the survey instrument (see Appendix I for details).

Table II.A: Total responses for:
I. Policy on Use/Wait

	Wait for				
Find ..	USE	CIP	Member	LC	:N/A:
CIP	99	0	1	7	: 0 :
Member	90	6	0	17	: :
LC	102	0	0	0	: :

Remark A: All libraries use LC, and all but 3 use CIP.

Table II.B: Total responses for:
II. Call number verification

			Only	
Copy	YES	NO	for some	:N/A:
LC	38	40	28	: 1 :
CIP	37	38	28	: :
Member	75	11	14	: :

Remark B: Note that the order of the rows in the table is different from that in Question 1. The most prominent response is 75 libraries verifying call numbers for Member copy. However, substantial numbers of libraries verify for LC and CIP as well.

Table II.C: Total responses for:
III. Policy on Revision

	Same	Different	N/A
LC-CIP	88	9	4
LC-Member	58	38	
CIP-Member	57	36	

Remark C: The largest similarity is LC-CIP, but it is not universal.

Table II.D: Total responses for:
IV. Authority Work

	Same	Differ	Different for some	N/A
LC-CIP	87	4	14	0
LC-Member	41	37	26	
CIP-Member	42	34	23	

Remark D: Again LC-CIP is the largest similarity, but not dominant.

Table II.E: Total responses for:
V. Categories of staff performing copy cataloging

	Same: pepl	Diff Ppl at same: level	Diffn levels	Profesl YES	Profesl NO	If yes: LC	If yes: CIP	If yes: Member	N/A
LC-CIP	88	5	17	40	61	13	13	38	0
LC-Member	54	12	46						
CIP-Member	54	14	42						

Remark E: This table summarizes the responses to several questions. Once again the LC-CIP similarity comes out. When professionals are used, they are primarily used to do Member-based copy cataloging.

3. Consistency Problems. A New Coding of the Data.

In Appendix II we present the complete data for each reporting library, in the form of several small tables. The reader will easily spot cases in which a particular option was marked "YES" and "NO" and "SOMETIMES."

To cope with this inconsistency we imposed a few logical rules. If a library reported that they wait for LC copy and Member or CIP copy, we scored them as waiting for LC

copy. If they reported that authority work is both "the same" and "different" for two kinds of copy, we scored it as being sometimes different. In question V we reduced the answers to just two categories: "same people" and "not the same people."

With this done, the responses of each library can be summarized in a series of revised tableaux, as shown in Appendix III. In addition, we scored each tableau for the degree of similarity it represents.

Meaning of the Values of S1,...,S5

5 = LC, CIP and Member copy treated the same;
4 = Only Member copy treated differently;
3 = Only LC copy treated differently ;
2 = Only CIP copy treated differently;
1 = All three classes treated differently from each other

The results of this cleaning and recoding are summarized in Appendix III. A typical case is shown in Exhibit 1.

Exhibit 1: Example of a revised tableau

```
TYPICAL*RO                1000   010   01   100   10
RO          110 34   10 1000   010   01   100   01
                      1000 0 100 0 01 0 100 0 01 1000101   5415 4 4
```

The summary data are the numbers 5415 4 4.

The systems used are RLIN and OCLC, leading to "110" in the "OCLC, RLIN, Neither" field. The library's Rank is 34 (by size measured by volumes added). Its serial number in the data file is 10. The first Tableau shows that all three kinds of copy receive the same treatment (they are used immediately), leading to the first "5" in the summary field. The next shows that call number verification is different for only Member copy, leading to the "4". The third shows that revision practice is different for all pairs, leading to a "1". Yet headings and authority work is the same for all types of copy, leading to a "5".

The last tableau, which refers to staffing, is separated in this display. Here it is the same only for LC and CIP, from which we conclude that for Member it is different, scored as "4". The next string of digits reports on the use of professionals.

1000101 = 10 001 0 1
10 means professionals are used in copy cataloging
001 means they are used only for full standard Member

This is scored as the final "4"

4. Analysis of the Data in Recoded Form.

We refer to the summary numbers, for short, as S1,...,S6. Their distributions are shown in Table III. Recall that the aspects are defined as:

[1] = Waiting for better copy
[2] = Verification of call numbers
[3] = Revision practice
[4] = Headings verification and authority work
[5] = Staff involved

Table III: Frequencies of the summary numbers S1,...,S6

Value	S1	S2	S3	S4	S5	S6
0	0	0	0	2	1	0
1	1	1	10	16	15	64
2	3	2	2	1	1	1
3	3	0	2	2	5	1
4	16	40	34	44	37	24
5	79	59	54	37	43	12
Total	102	102	102	102	102	102

The definition of S1,...,S5 are given on the previous page. The value (0) indicates a missing value or a logical inconsistency.

For S6 the meaning of the codes is.:
1 Professionals do not do any cataloging
2 Professionals do LC cataloging
3 Professionals do CIP cataloging
4 Professionals do Member cataloging
5 Professionals do all cataloging

We see that most libraries fall into either group "4" or group "5." In fact, hope springs at once that perhaps as many as 37 of the libraries will show the pattern "55555" and another 16 will show "44444." Although this is logically possible, given the observed frequencies, it just doesn't happen. The entire frequency distribution for the 5-number overall description is shown in Appendix IV.

There are only a few cases which occur more than once in that distribution. They are

summarized here in Table IV.

Table IV: Partial frequency table of the 5-number patterns

CASE	FREQUENCY
55555	11
54555	5
55554	4
55545	4
54444	4
55544	3
55541	3
44444	3
55551	2
55511	2
55445	2
55444	2
55145	2
54545	2
54454	2
54445	2
54411	2
45545	2
44414	2

We notice that these patterns are composed only of "5"s "4"s and "1"s: all the same, Member different, or none the same. Most of them occur only twice. The only ones that occur more than twice are made all of "4"s and "5"s. That is, at most Member is treated differently. The most common pattern is to treat all entirely alike "55555" and yet that is seen at only 11 of 102 libraries. The next most common pattern is to treat Member differently, but only for call number verification, then for staffing, and then for headings verification. Combined, these categories account for 24 of the 102 libraries. A very weak plurality at best.

Further analysis of this kind (including a rather sophisticated test of whether any of these variables could, taken together, explain the rest) led to no additional insight. The most promising summary description is simply to say that:

At more than half the libraries, practice with regard to each of the 5 aspects of copy cataloging either treats all copy in the same way, or treats only Member copy differently. In our code, such a library is represented by only "4"s and "5"s.

We may summarize those patterns in Table V.

Table V: Number of "4-5" patterns

All 5's	11 Cases
4 5's and 1 4	14
3 5's and 2 4's	11
2 5's and 3 4's	10
1 5's and 4 4's	5
All 4's	3
TOTAL	54

This concentration ignores the fact that there are many different patterns of, for example, 3 5's and 2 4's, according to which activities are different for Member copy. Thus this summary includes many singletons (patterns found only at one library.)

5. The Search for Explanations of the Patterns.

Since, conceivably, the staffing patterns play a strong role in determining the policies (or vice versa), we looked at the breakdown of the first four codes by the value of the fifth and sixth codes. The detailed tables, summarized in Appendix V, showed no significant features. In other words, knowledge of the staffing assignments did not predict the policies (or, presumably, vice versa).

It was considered that the maturity of the local on-line system might have some effect on policies. This was checked by extracting the 10 libraries judged (by Carol Mandel) to have relatively mature local systems. The distribution is undistinguished. The sample is so small that we would expect at most one occurrence of 55555 and in fact, we got none.

Table VI: Frequency table of the 5-number patterns for "mature" libraries

MATURITY=1

S1S2S3S4S5

Value	Freq
44554	1
54543	1
55443	1
55505	1
55511	1
55543	1
55545	2
55551	1
Total	9

In short, we did not find that the patterns of staffing, or the maturity of the local system, is strongly linked to the patterns of copy cataloging. Certainly we could not say that the patterns of staffing "explain" the other patterns that have been observed.

6. The Effects of Library Size.

A third variable which might affect patterns of policies for copy cataloging is the size of the library's cataloging effort. As a surrogate for this characteristic we used the RANK of the library, as measured by volumes added, reported in the ARL Statistics for 1985-86. We divided the libraries into three classes: BIG=1 = {libraries with RANK 1 to 32}; BIG=2 = {libraries with RANK 33 to 67}; BIG=3 = {libraries with above 67}. We determined the crosstabulation of the policy variables S1,...S6 with BIG. The only significant correlation is in variable S6. The results are summarized in Table VII.

Table VII. Use of Professionals in Copy Cataloging.

BIG =	1	2	3
Don't use professionals	24	17	20
Do use professionals	6	12	15
	30	29	35

It is clear that the largest libraries are far less likely to use professionals in copy cataloging. Further analysis shows that at the 6 large libraries where professionals do copy cataloging, they handle only Member copy. At the smaller ARL libraries, in 1/3rd of the cases, professionals handle all types of copy. This difference may be the result of the larger work flow, which makes it possible to divide copy cataloging into a number of routine streams. On the other hand, it may represent the persistence, in smaller ARL libraries, of professional staff in tasks that could be assigned to non-professional staff. Detailed investigation of the cases would be needed to resolve this question.

The specific patterns of policy also show some dependence on the BIG variable. For the largest libraries the most common pattern for the first four aspects is 5554. (Authority for Member copy is different.) But when all five aspects are considered this pattern breaks into several different staffing patterns. On the other hand, the 55555 pattern, which is the only clear leader in the analysis of all libraries, is significantly absent at large libraries, as shown in Table VIII.

Table VIII. Breakdown of libraries treating all aspects of copy cataloging in the same way.

	BIG =	1	2	3
Pattern 55555		1	4	6
		30	29	35

As with the staffing patterns, this difference may reflect the potential for dividing a larger work flow into several streams. It may also reflect a heightened perception of the need to monitor the work of other libraries and/or the need to maintain the integrity of local authority files.

7. Summary.

The 102 libraries reporting in this survey exhibit a bewildering variety of patterns of behavior, with regard to copy cataloging. The only pattern found at an appreciable number of libraries is to "treat everything the same." This pattern is found, however, at only 10% of the libraries. Thus it can hardly be called dominant. It is less common at large libraries.

We find that the most common deviation from this pattern is to treat only Member-based copy cataloging differently. When this deviation is included, a total of 54 libraries are accounted for. We find that this number is about what would be expected if there were no correlation among the five aspects of policy studied here. In other words, if libraries selected their policy for each of the five aspects without regard to the rest of their policies, the observed distribution would arise. This does not mean that libraries set policy at random. But it does mean that the statistical analysis alone cannot expose the reasoning behind the selection of policies.

It is of great interest to speculate on whether the choice of a pattern of policies shows significant correlation with the costs of copy cataloging. In general, it is hard to compare costs across libraries. However, the relation between cost of cataloging from LC/CIP copy can be compared with the cost of cataloging from Member copy at the same library. Such a comparison might show some effect of the pattern of policies, but would require data which are beyond the scope of this study.

8. Acknowledgements and Responsibilities.

The survey instrument was designed by Carol Mandel, chair of the BSSC, in consultation with P. Kantor. The survey was administered by Ellen Timmer at the Council on Library Resources, who entered the data into machine readable files. The data were analyzed by M. Cherikh, using the ABSTAT statistics program. S. Rich provided the codes to create Appendices II and III, and Table V. The plan of analysis was developed by P. Kantor, who authored this report.

COSTS AND COST BENEFITS OF DISTRIBUTED CATALOGING TO LIBRARY OF CONGRESS STANDARDS

COSTS AND COST BENEFITS OF DISTRIBUTED CATALOGING TO LIBRARY OF CONGRESS STANDARDS

Paul B. Kantor, PhD

EXECUTIVE SUMMARY

1. THE DIRECT COST BENEFIT MODEL.

We make several working assumptions to reduce the complexity of distributed cataloging to a model with just a few key parameters. These parameters are defined as:

a = The percent cost increase in cataloging to NCCP standards.
b = The percent decrease in deriving from LC/CIP/NCCP records.
c = The average cost of creating an Ordinary Original record.
d = The average cost of a derive from Ordinary Original record.

Each time a book is cataloged to NCCP standards as opposed to Ordinary Original Cataloging (OOC) there is an added expense **a** x **c**, at the cataloging library. For example, if **c** = $50.00 and **a** = 75% then the added cost is 75% of $50 = $37.50. Similarly, the cost saved (at a different library, of course) is **b** x **d**. [If **d** = $20.00 and **b** = 40% this is $8.00]. For realistic numbers, the savings is much smaller than the added expense.

The two cost studies reported here have determined two of these four parameters:

a = 75%.

b = 37%.

2. The more complete model, now under development for the BSSC, includes a computation of the breakeven holdings level necessary for the cost savings to cover the added cost of cataloging to the NCCP standards. It also includes the subtle "reflected effect" in which LC uses NCCP copy to free resources for the creation of additional LC records. This has the effect of reducing the breakeven holdings level.

DETERMINATION OF THE ADDED COST OF NCCP CATALOGING: "a"

PURPOSE

This study was undertaken to estimate, by studying a sample of titles cataloged at the NCCP libraries, the difference in cost between Ordinary Original Cataloging at those libraries, and cataloging to the LC standards, using the NCCP procedures. To improve comparability among libraries with differing pay scales, the result is expressed simply as the percentage increase coefficient "a".

DETERMINATION OF "a": SELECTION OF SAMPLE

Contraints of Data Collection. Any cost study poses a burden to the workers who are being studied. This motivates making the sample of items studied as small as possible. On the other hand, there are certainly individual variations in the time required to catalog specific items. The sample should be large enough to iron out these variations. While statistical theory favors the use of a truly random sample, the drawing of such a sample requires substantial intervention in the daily routine of the workers, and causes the study to extend over a long calendar period.

Collection Methods. As a compromise among all of these goals, NCCP libraries were offered their choice of two data collection methods. The first, called the Work Slip method, used a data collection form which travelled with the book from the beginning to the end of the cataloging process [Exhibit 1], and on which each worker noted the number of minutes required for the several activities. The second, called the Spreadsheet Method, used tabular data collection forms on which workers recorded, day by day, the total time spent at each of several activities, and the total number of items completing that specific activity on that day. [Exhibit 2].

Sample Size. The size of the recommended sample for these two cases was not the same. For libraries using the work slip method we suggested a sample of 100 ordinary events and 100 NCCP events. Only one library had sufficient work flow to reach these levels in a timely fashion. As may be expected, the use of work slips may greatly prolong the study, as each item makes its way through the pipeline. At one library, problems in telecommunication caused substantial delays of this type. For libraries using the Spreadsheet Method, we recommended that data be collected continuously for a 4-week period. We estimated that three weeks of data would be sufficient, and the use of a 4-week period made it unecessary to schedule around holidays, professional meetings, and other interruptions.

The Control. For each library, in addition to the person(s) doing NCCP cataloging, the study must designate some person(s) who will do Ordinary Original Cataloging to serve as the base line for comparison. Since the productivitiy of catalogers is rather variable, and depends on external factors such as the language of the books cataloged, the assignment of controls was a troublesome feature of the experimental design. In some case the control included the same person(s), who did both types of work. In other cases it was other

catalogers, selected by the local Project Manager as being approximately comparable. In some cases it was not possible to match the language. All of these problems introduce systematic uncertainties in the final result, whose magnitude is likely to be comparable to the statistical uncertainties reported below.

DETERMINATION OF "a": PROCEDURES

At each library, contact was made through the cognizant member of the NCCP Operations Committee (generally the Head of Technical Services) who, in turn, designated a point of contact to serve as Project Manager (PM). After discussions with Tantalus, the PM selected the NCCP cataloger(s) to participate in the study, and the control(s). The data collection instruments were refined by pre-testing at the University of Chicago and at Harvard University. The PM and staff at each library were free to choose between the two instruments. Libraries using the Work Slip method were permitted to alter the details of the work slip to correspond to local usage, subject to approval by Tantalus. Several units adopting the Spreadsheet Method found it useful to add further columns representing local practice. One library actually developed a parallel set of data collection forms to reflect the fact that Authority Work is carried out, at that library, in a distinct unit.

Data were collected early in 1989. Two of the libraries agreed to repeat the data collection, and did so in early Fall 1989. At the time of the data collection only one of the libraries had been designated "Independent" with regard to every aspect of cataloging. Data were entered into specially designed spreadsheets for analysis. Results were transmitted to the participating libraries for discussion and comment. In some cases the method of analysis was revised, to more accurately reflect local practices. Libraries were asked to speculate on possible explanations for data which fell far from the center of the range.

In essence, the data from each library were treated as follows. The time spent by each worker was multiplied by a nominal salary per minute (based on the assumption that the annual salary represents 2,000 hours). The total nominal salary cost of all workers engaged in a particular activity was divided by the total number of items completing that activity during the study. The resulting nominal unit costs per activity were summed over the specific activities (Cataloging, Editing, Input, etc) to yield a nominal unit cost for complete cataloging. Finally, the cost increase coefficient "a" was determined as:

a = (Nominal NCCP Unit Cost)/(Nominal Ordinary Unit Cost) - 1

Nominal Costs are not reported here, but were reported to the individual libraries, with the suggestion that actual costs, including fringe and "non-productive time", are likely to be double the nominal costs.

DETERMINATION OF "a": RESULTS FROM NCCP LIBRARIES.

The results of this study are summarized in Table 1. The libraries are coded LA to LG in order of increasing value of the increase parameter **a**, to preserve anonymity. The

measured value ranges from 25% through 151%. The number of items on which this estimate is based is reported for each library. In all 7 studies together, 567 Ordinary items and 461 of NCCP items were studied. In almost all cases where data are avilable, the NCCP work was done from "scratch" or from nothing more than the LC in-process file entry ("APIF"). There were 7 items processed from initial Minimal Level Records.

TABLE 1: FINAL DATA SUMMARY FOR NCCP CATALOGING COSTS: LIBRARY DATA

	Incrs a	Number of items		Originating Records		
CODE	(Percent)	OrdOrig	NCCP	Sctch	APIF	MLC
LA	25%	95	53	32	21	
LB	40%	100	101	91	10	
LC	46%	59	67			
LD	73%	58	84			
LE	74%	167	54	54		
LF	84%	50	49	39	3	7
LG	151%	38	53	37	16	

Source: [ds-105\clr\nccp\table1.wrk 89-12-21 14:46]
Notes to Table 1
The first column shows the library code. The second shows the percent increase in cost when cataloging to the NCCP standards. The third and fourth columns report the number of items studied. The remaining columns indicate the source or basis for the NCCP records created.

LA. This library's analysis was done using the spreadsheet method rather than work slips. The librarians there report that none of the work is done from Member copy and approximately 60% was done from scratch and 40% from the LC in-process file. It is significant to note that, at this library the cataloger does inputting directly for NCCP, while for ordinary original cataloging there is a separate inputter, requiring revision. There are two NCCP catalogers and they revise [that is, check] each others work after the inputting. The control cataloger is not one of the NCCP catalogers. The catalogers involved report good (human) communication on a friendly basis with the Library of Congress.

LB. This library gathered data using the work slip method. There are no significant exceptions to be noted.

LC. This library gathered data using the spreadsheet method. After careful review in January 1990, all data were treated by the aggregated method. This did not change the earlier results.

LD. This library, which used the spreadsheet method, represents the median value that will be used to carry forward the economic analysis.

LE. This library was one of several that offered to do a complete second data collection, because of possible problems with the first data collection. There have been changes in policy between the two data collections which result in significantly more time spent on the NCCP work. The revised figure for the percentage increase is in the center of the pack.

LF. This library is quite close to the median. The data for this library were gathered by the work slip method. A detailed account of communication costs was reported. This library reported that 7 of the NCCP titles were based on Minimal Level catalog records.

LG. This library, which used the work slip method, is the outlier at the high end. This library reports that "a large portion of the original cataloging is brief record cataloging." It also reports time spent in verification and migration from one system to another.

The original design called for separating the observed cost increase into the several cataloging activities (Cataloging, Editing, ...) but the interim data analysis revealed that categories were not strictly compatible at the several libraries, and the data showed uninterpretable variation. For this reason, the corresponding breakdowns are not reported here. Related to this situation, we have had to use best judgment in a number of cases, to impute the total number of items to which a specific cost must be applied. At one limit we have the full aggregated estimate: the total of costs is applied to the total number of items processed during the sample period. At the other extreme we have full disaggregation: the number of items for which any process was studied may be different but it is assumed that each such process must be done for every item cataloged. The choice between these two approaches can significantly affect the nominal costs, and the increase "a" reported at any particular library. We have advised each library of this problem and made every effort to find the "correct denominator" for the unit cost calculations. We have observed, during revisions, that the general range of results seems fairly stable against these changes in the assumptions and interpretation.

DETERMINATION OF "a": CONCLUSIONS

The immediate conclusion is that, at these seven libraries, for the samples of items studied, there is enormous variation in the percentage increase in costs when NCCP cataloging is done. This variation, a factor of 6, is large even for library economic studies. The underlying nominal cost data (not reported here) showed almost 3-fold variation for the cost of Ordinary Original Cataloging and almost 4-fold variation for the cost of NCCP cataloging. But much of this variation is attributable to differences in wage scale and local work practices.

Discussions with all of the libraries reveal that communication costs contribute substantially to the increase, and it may be that those costs decrease when steady state is reached. The highest figure (151%; Adjusted Value 153%) was measured at a library which reports that much of the control material is Brief Record Cataloging which is less expensive than full Ordinary Original Cataloging. This would, of course, make NCCP relatively more expensive.

To carry forward the combined economic model we must extract from Table 1A a representative value for "a", and some indication of the range in which "a" might lie if it were to be measured at every library in the ARL (should the NCCP be extended to that range.) The most natural choice for a single value is the median or mid-point, 75%. The range may be taken as 43% to 88%, which includes all but the two extreme cases.

[In terms of purely statistical confidence limits, the chance that this range lies above the true population median, if these seven are regarded as drawn at random from such a population, is approximately 10%. Simlarly, there is approximately 10% chance that this range is too high. Thus it is an 80% confidence interval. Such an interval is usually not regarded as statistically "persuasive."]

In addition to the relatively weak statistical confidence of this estimated interval, we must recall that systematic errors, including variations in the choice of the control person(s), add further uncertainty. The representative value, a = 75%, must therefore be regarded as indicative but not at all determinative.

From a management point of view it is always interesting to examine the extremes. We have already discussed the likely reason for the upper extreme. Of greater interest is the lower extreme. As shown in the detailed notes to Table 1, this library appears to have established a comfortable and collegial relationship with personnel at LC. In addition, the work mode chosen for NCCP cataloging eliminated some costs associated with ordinary cataloging. This change could be used to reduce costs at any library. The establishment of comfortable working relations might also be viewed as an economically desirable management goal.

DETERMINATION OF THE COST SAVINGS WHEN NCCP RECORDS ARE AVAILABLE "b"

PURPOSE

This study was undertaken to estimate, at a sample of ARL libraries, the benefit coefficient "b" representing the fractional decease in cost when copy cataloging is based upon LC or CIP rather than Ordinary Original records. Our model assumes that the same benefit will be experienced in cataloging from NCCP records created to the LC standard.

SELECTION OF SAMPLE

The sample was selected by a rather complex procedure whose value, after the fact, became dubious. Our working assumption was that the size of "b" would depend in some way on the degree to which the two types of copy cataloging [based upon LC or Cataloging in Publication (CIP) records *vs* based upon Ordinary Original (also called "member") records] followed the same procedures at a given library.

To this end a survey of the ARL libraries was carried out, determining the patterns of policy with regard to five key aspects of derived cataloging. A preliminary analysis of this data led to a ranking into nine distinct categories. At one extreme were libraries for which the policies either treated these two kinds of derived cataloging in typical and similar ways, or treated them in atypical ways. At the other extreme were those which either treated them in typical but differing ways, or treated them in atypical ways. The intent was to delineate a spectrum from libraries which had a typically uniform policy to libraries which had a typically differentiating policy. Uniform policy, it was thought, would correspond to small values of "b"; differentiating policies would correspond to large values of "b." All 102 libraries responded, and more than 60 indicated willingness to be among the sample libraries for the cost study.

Details of the sample selection have been reported elsewhere. A sample was

generated randomly, subject to the requirement that each of the nine "policy classes" as then defined be represented, that the utilities be represented in proportion, and that the libraries be reasonably distirbuted with regard to size (as measured by the ARL Volumes Held statistic.) A few changes to this random sample were made "by hand" to improve the large library representation. One of the selected libraries found that it could not resolve certain costs at the level called for in the analysis, and was replaced by another library in the same policy class.

We recommended that data be collected continuously for a 4-week period. We estimated that three weeks of data would be sufficient, and the use of a 4-week period made it unecessary to schedule around holidays, professional meetings, and other interruptions.

PROCEDURES

At each library contact was made through the library director, who in turn designated a point of contact to serve as Project Manager (PM). After discussions with Tantalus, the PM selected the copy catalogers to participate in the study. The data collection instruments were based on the Spreadsheet form pre-tested at the University of Chicago.

Data were collected early in 1989. One of the libraries based its report upon data collected during a previous internal study of the same question. Data were entered into specially designed spreadsheets for analysis. Results were transmitted to the participating libraries for discussion and comment.

In essence, the data from each library were treated as follows. The time spent by each worker was multiplied by a nominal salary per minute (based on the assumption that the annual salary represents 2,000 hours). The total nominal salary cost of all workers engaged in a particular activity was divided by the total number of items completing that activity during the study. The resulting nominal unit costs per activity were summed over the specific activiities (Cataloging, Editing, Input, etc) to yield a nominal unit cost for complete cataloging. Finally, the cost decrease coefficient "b" was determined as:

b=1-(Nominal LC/CIP-based Unit Cost)/(Nominal Member-based Unit Cost)

Nominal Costs are not reported here, but were reported to the individual libraries, with the suggestion that actual costs, including fringe and "non-productive time" are likely to be double the nominal costs.

RESULTS

The results are summarized in Table 2

TABLE 2: FINAL DATA SUMMARY FOR COPY CATALOGING COSTS

CODE	Reductn b (Percent)	Number of items LC-base	CIP-base	Mmbr-base
CA	-49%	284	423	448
CB	-5%	733	352	514
CC	19%	2023	50	1035
CD	21%	1042	355	640
CE	30%	3522	133	1648
CF	34%	1554	330	734
CG	40%	1908	1270	1015
CH	45%	1548	1163	325
CI	50%	735	752	626
CJ	51%	807	345	927
CK	68%	1102	268	443
	TOTALS	15,258	5,441	8,355

Source: [ds-015\clr\copy\table2.wrk 89-12-21 16:24]

The first column gives the library code. The second gives the decrease coefficient "b". The remaining three columns report the number of records of each type that were processed during the sample period.

Code Notes to Table 2

CA LC/CIP initiates other work at this library
CB We do not know why this is negative
CC No authority work is done at all
CD Time of original catalogers is added in here (revision)
CE Only Auth work aggregated
CF Data adjusted to include professionals
CG Auth work, for reports only, dropped
CH Auth work reported by distinct unit
CI Auth work aggregated
CJ Professionals heavily involved in member copy
CK Auth work aggregated

The original design called for separating the observed cost increase into the several cataloging activities (Cataloging, Editing, ...) but the interim data analysis revealed that categories were not strictly compatible at the several libraries, and the data showed uninterpretable variation. For this reason, the corresponding breakdowns are not reported here. Related to this situation, we have had to use best judgment in a number of cases, to impute the total number of items to which a specific cost must be applied. At one limit we have the full aggregated estimate: the total of costs is applied to the total number of items processed during the sample period. At the other extreme we have full disaggregation: the number of items for which any process was studied may be different but it is assumed that each such process must be done for every item cataloged. The choice between these two approaches can significantly affect the nominal costs, and decrease "b" reported at any particular library. We have advised each library of this problem and made every effort to find the "correct denominator" for the unit cost calculations. We have observed, during revisions, that the general range of results seems fairly stable against these changes in the assumptions and interpretation.

CONCLUSIONS

The immediate conclusion to be drawn from this table is that there is substantial variation in the cost decrease achieved by cataloging from LC/CIP records, compared to cataloging from Ordinary Original records. Further, as with most library economic parameters, we see substantial variation. The large negative number at library CA (which means that cataloging from LC/CIP base is MORE expensive) may be disregarded because it results from that library's special policies, as noted. When they work from LC/CIP they do additional authority work which is not done for other derived cataloging. Library CA regards this as a temporary expedient, pending the adoption of an automated system for authority control. The second negative value has not been explained, but, given the size of the sample and the precision of the methods, it is consistent with the value 0.

Except for case CA, then, the numbers vary from a low of essentially 0% (no difference and no savings) to a high of 68% (more than two-thirds savings). We have explored these data from a number of perspectives, including reaggregating the data from specific libraries in other ways, and the essential variation remains unchanged. Hence the selection of a representative value for the remainder of the analysis is somewhat risky. The central numbers in this list are 34% and 40%. Their average can be taken as a representative midpoint: 37%. A range excluding the lowest and highest points runs from 19% to 51% and could be adopted as an interval estimate of the population value of the savings parameter "b".

[Statistical significance is somewhat higher for this situation because the sample is larger. The chance that the indicated range would lie above the true population median is only 1%, while the chance that it would lie below the population median is also 1%. However, the range itself represents a factor of 2.5, so the price for our improved confidence is substantial imprecision. As the numbers stand, there is no point in reducing the interval, because there is little change in the endpoint values at reduced confidence.]